From Shipping to Swimming:

Bacterial Monitoring and Diversity in Ballast Treatment Systems and Recreational Waters

A Thesis

SUBMITTED TO THE FACULTY OF THE

UNIVERSITY OF MINNESOTA

BY

Lisa DeGuire

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF MASTER OF SCIENCE

Advisor: Dr. Randall Hicks, Dr. Chan Lan Chun

August 2021

Acknowledgments

Throughout this research I have received a great deal of help and support.

I would first like to thank and acknowledge my advisors Dr. Randall Hicks and Dr. Chan Lan Chun. Their mentorship, expertise, time and feedback have been invaluable throughout my graduate experience. Thank you also to Dr. Chris Filstrup for serving on my thesis committee and NRRI's Analytical Lab sharing their space and lab equipment throughout sample processing.

Thank you to fellow members of the Hick's lab Julia Zimmer, for her comradery and lab support and Jennifer Knack for her assistance processing sequence data. I would also like to thank the members of the Chun lab who helped to make this research possible. In particular, I would like to thank Miranda Galey who was an essential bioinformatics resource, performed the SourceTracker analysis, and contributed the library-dependent (SourceTracker) section of the methods. Thank you also to Max Brubaker and lab assistants Collin Krochalk, Brendon Gearhart, Joseph Pesik, and Emilie Ziebarth for their help processing water samples, extracting DNA and ion analyses. Thanks also to fellow lab mates Shawnee McMillian for her assistance with the ion analyses and Adelle Keppers for being such a helpful lab resource and thesis accountability partner.

Thank you to the staff at the Lake Superior Research Institute for their help implementing the ballast treatment experiment, including Matt TenEyck, Heidi Saillard, Christine Polkinghorne, Kelsey Prihoda, Kimberly Beesley, Lana Fanberg, and Alexander Frie.

I would like to acknowledge the staff at the Lake County Soil and Water Conservation District for their help sampling and transporting and helping to enable the microbial source tracking project. These staff include Derrick Passe, Emily Nelson, Zach Stewart and GreenCorps intern Macy Osborne.

Thank you also to the University of Minnesota Genomics Center for performing the DNA sequencing for this research.

Additionally, I would like to thank the following organizations for funding this research. Funding for the ballast water treatment portion of this project was provided by the Minnesota Environment and Natural Resources Trust Fund (project M.L. 2016, Chp. 186, Sec. 2, Subd. 06c) as recommended by the Legislative-Citizen Commission on Minnesota Resources (LCCMR). Minnesota's Lake Superior Coastal Program helped to provide financial support for the stream microbial source tracking project. Additional support came from the University of Minnesota Duluth Biology Department, and the University of Minnesota Water Resources Science Graduate Program.

And finally, thank you to my amazing parents and friends for their constant love and support throughout this journey.

Table of Contents

List of Tables

List of Figures

Background:

The use of indicator organisms as a proxy for pathogenic bacteria significantly reduces the cost and complexity of monitoring aquatic systems. Fecal indicator bacteria, such as *Escherichia coli* and *Enterococcus spp.*, are commonly used by governmental agencies in recreational beach monitoring, water and wastewater treatment, and more recently as a tool in ballast water management (IMO n.d).

Indicator microorganisms are generally selected due to their (a) well established and cost-effective monitoring methods, (b) co-occurrence with other harmful pathogens, and (c) abundance being generally higher than specific pathogens (Grabow 1986). *E. coli* and *Enterococcus* spp. are bacteria that primarily live in the gastrointestinal tracts of warm-blooded animals and therefore are often used to indicate fecal contamination. While both genera contain pathogenic and non-pathogenic strains, it is not usually the indicators that pose a public health risk, but rather more dangerous pathogens that may also be present in fecal inputs. Due to their ability to survive in salt water, *Enterococcus* spp. are the recommended indicator organisms for saline and brackish waters while *E. coli* is more commonly used in freshwater monitoring systems (USEPA n.d.).

Historically, the abundance of indicator bacteria has been determined through culture-based methods (USEPA 2014). Culture-based quantification involves transferring a water sample to a growth medium, incubating and counting the cultures (or colony forming units) that appear. Molecular methods, such as quantitative polymerase chain reactions (qPCR), can be used for the same purpose (to indicate and quantify fecal

contamination); however, the underlying concept is quite different. qPCR involves amplification of a DNA segment from a target organism and quantifying its relative abundance through the detection of a fluorescent signal directly proportional to the amplified DNA. While culturing involves the growth of actively reproducing bacteria, qPCR amplifies any target DNA present regardless of an organism's viability. Because the techniques measure different metrics, the results of these two approaches may not always be correlated.

There are benefits and challenges to both techniques. Cultivation is relatively simple, user-friendly, and measures actively growing cells, but only selects for media-specific bacteria capable of growth on liquid or semi-solid media and can require long periods of incubation. Additionally, culture-based approaches can underestimate bacterial abundance as they do not account for cells that are viable, but non-culturable (VBNC) (Ahmed et al. 2015). This VBNC state is a microbial survival strategy which arises in response to environmental stress and adverse conditions. As their name implies, these cells are not able to be cultured, but they retain their viability and virulence. qPCR is more precise, can be designed to detect a wide range of target organisms, and easily paired with other molecular techniques like next generation sequencing (NGS). Both qPCR and NGS can have high startup costs, though, and do not indicate viability of the cells that contributed the target DNA. Some recreational monitoring programs have recently adopted qPCR over culturing because it can provide results within hours of sample collection (once laboratory protocol and infrastructure are in place) instead of requiring overnight culture incubation and therefore delayed advisories (Aw et al. 2019).

This research utilized molecular techniques to explore the use of indicator bacteria as a measure of water quality in two projects: (1) a bench-scale experiment to explore the effectiveness of ballast water treatment techniques in freshwater and (2) an investigation of sources of fecal contamination of the Skunk Creek watershed in Two Harbors, MN.

The overarching objectives of this research were to (a) investigate the use of indicator bacteria (*E. coli* and *Enterococcus sp.)* as measures of water quality in ballast treatment systems and stream monitoring and (b) compare culture-based and molecular techniques as tools for bacterial monitoring. Both projects explored the relative abundance of indicator bacteria and overall bacterial community composition generated through molecular (qPCR and DNA sequencing) and culture-based methods (IDEXX QuantiTray).

Ideally, the results of this research will inform future monitoring and treatment practices and subsequently benefit regional water quality. The goal of the ballast experiment was to provide information on whether established indicator bacteria are truly representative of other potentially harmful microbes, as well as the scale of post-treatment microbial regrowth and, ultimately, reduce the risk of introducing harmful bacteria into local waterways. Identifying the source of elevated *E. coli* in Skunk Creek and Agate Bay should enable Lake County and the city of Two Harbors, MN to better understand the extent of human health risks and take appropriate action to mitigate future inputs.

Chapter 1: A Bench-scale Evaluation of Ballast Water Treatment for Bacterial Communities

Summary

Ballast water is a major vector for the spread of invasive species, and the Duluth-Superior Harbor (DSH) receives the most ballast water discharge of any port in the Laurentian Great Lakes. While most concern has focused on plant and animal invasive species, potentially harmful microbes have been largely overlooked. A bench-scale experiment was conducted to evaluate whether indicator bacteria are truly representative of other potential pathogens after ballast treatment, as well as the scale of post-treatment bacterial regrowth. Two common treatment techniques (UV light and chlorination) were performed on ambient water collected alongside a lake freighter in the DSH, half of which was spiked with indicator bacteria. Culture-based quantification, qPCR, and 16S rRNA amplicon sequencing were performed on samples immediately after treatment, as well as five days after treatment to check for bacterial regrowth. Both treatment techniques resulted in 99% reductions in culturable indicator and heterotrophic bacteria immediately following treatment (68-99% reductions when measured by qPCR). After 5 days, however, both lab and field incubations showed considerable regrowth of total bacteria (not reflected in indicators) and a distinct shift in bacterial community composition, including the regrowth of multiple pathogen containing genera (particularly *Acinetobacter, Flavobacterium, Pseudomonas*). These results can be used to inform ballast management decision makers as they assess the risks and treatment options regarding potentially harmful microbes.

Introduction

The rise in international trade over the past century has brought with it the introduction of many invasive species to the Great Lakes ecosystems. Ballast water has been identified as a primary vector for the spread of invasive species (EPA 2011). Duluth, Minnesota has been labeled an invasion "hotspot," as it receives the most ballast discharge of any Great Lakes port (Duluth Seaway Port Authority n.d.). Each year, Duluth-Superior Harbor (DSH) supports an average of 900 visits from bulk cargo vessels. These visits include both "lakers," which travel within the Great Lakes, and ocean vessels known as "salties." Due to heavy shipping traffic and discharge within the DSH, many of the major Great Lakes invaders including gobies, ruffe, zebra mussels, and quagga mussels have been detected in the St. Louis River estuary ecosystem (Minnesota Sea Grant 2017). Much less is known about native or invasive microbial populations and communities in the DSH.

Invasive microorganisms have been overlooked and understudied largely due to the challenges associated with detection and monitoring (Litchman 2010). While many microorganisms have a beneficial or neutral impact on humans and ecosystems, invasive and pathogenic microbes can pose a threat to ecological and human health. One example is the bacterial pathogen *Piscirickettsia salmonis* (known to cause "Muskie pox") which infects Salmonid and muskellunge fish populations causing up to 90% mortality (Marshall et al. 1998). Though small in size, microorganisms can have large impacts. Despite the logistical difficulties, microbial populations should not be overlooked in ballast water monitoring and regulation.

It is impossible to completely prevent the transfer of microorganisms from place to place as they are ubiquitous. Therefore, the issue becomes establishing an acceptable standard to lower the risk of microbial species invasion and pathogen introduction. Many pathogens can be difficult and expensive to culture and monitor, and often have patchy distributions or low concentrations in natural waters (Field & Samadpour 2007). Currently, the solution involves using indicator bacteria. While the use of indicator microorganisms as proxies for potential pathogens significantly cuts down on the cost and complexity of monitoring an environment, more research is needed regarding whether selected organisms are effective indicators of potentially harmful and invasive microbes within ballast water communities.

In 2004, the International Maritime Organization (IMO) adopted the International Convention for the Control and Management of Ships' Ballast Water and Sediments (BWM Convention). The BWM Convention requires that all ships implement a ballast water management (BWM) plan, carry a record book of all ballast activity, and adhere to the standards put forth in the Convention. Standard D-2 creates limits for "viable organisms" allowed in ballast discharge (including indicator bacteria). Standard D-3 states that ballast management system requirements must comply with the Convention and be approved by the IMO. Though adopted in 2004, the BWM Convention became effective on September 8, 2017. To date, 80 countries have ratified the BWM Convention (IMO 2019). Canada acceded to the Convention in 2010; the United States has not.

The United States Coast Guard (USCG) created the Ballast Water Management Act in 2005. One key difference between IMO and USCG regulations is that USCG

language limits the quantity of "living organisms" discharged in ballast water, rather than "viable organisms" as specified in IMO standards. The debate on the limits of "live" vs. "viable" microbes ended on November 27, 2018 with the passage of the Vessel Incidental Discharge Act (VIDA) within the USCG Authorization Act (EPA n.d.). This act was designed to reconcile the language between international, national and state-specific standards. This act included expanding the definition of "living" to exclude "nonviable" organisms. Because nonviable organisms are no longer considered to be live organisms, treatment methods that do not kill microorganisms, but render them incapable of reproduction, are acceptable techniques. VIDA also notably prohibits state and local authorities from enforcing and adopting regulations that differ from those of USCG. States are also prohibited from requiring permits for incidental discharge from small vessels (those less than 80 ft) and fishing vessels (of any length). While VIDA is largely considered a victory for maritime trade, regional opponents have demonstrated concern in treating the Great Lakes identical to all international seaways.

The guidelines for ballast treatment of microbial communities are based on discharge standards of the indicator organisms: *Escherichia coli*, *Enterococcus* spp. and toxigenic *Vibrio cholerae*. According to national and international standards, the abundance of indicator bacteria may not exceed 1 colony forming unit (CFU) per 100 ml for *Vibrio cholerae*, 250 CFU per 100 ml for *E. coli*, and 100 CFU per 100 ml for intestinal enterococci.

These indicator microorganisms are not always present in ballast water (Ng et al. 2015). In 2011-2012, twenty-one ballast tanks were sampled in DSH, and DNA from the

Escherichia and *Enterococcus* genera were not detected in any of the tanks based on 16S rRNA sequencing data (Knack et al. in prep.). However, the absence of indicator bacteria does not necessarily mean that harmful microorganisms are not present. It is also unclear whether the fates of indicator bacteria are representative of the whole bacterial community, especially given different treatment techniques, ballast water chemistry and intake/discharge environments.

Like the ships that contain them, ballast tanks come in all different styles, shapes and sizes and carry water from many different environments. Therefore, no one-size-fits-all approach can be applied to ballast treatment. Ballast management can be shipboard or port-based; physical, chemical, or mechanical; and can occur at various points throughout the intake/discharge process (Tsolaki & Diamadopoulos 2010). Popular mechanical treatment options include the application of UV radiation, heat, and deoxygenation, while common chemical options include the addition of chlorine, biocides and ozone. Over 48% of commercially available ballast treatment systems utilize UV radiation and 28% involve electrochemical treatment (Hess-Erga et al. 2019). Ballast water chemistry may have some impact on bacterial removal rate. For example, UV treatment on saltwater may be less effective as the scattering effect or light absorption by inorganic compounds may affect inactivation rates of the bacteria (Chen et al., 2016, Hess-Erga et al. 2019).

Both UV and electro-chlorination have exhibited immediate reductions in cultivable bacteria of up to 99% (Waite et al. 2003, Hess-Erga et al. 2019). These removal rates (99% for UV and chlorination) were also seen in culture counts of a preliminary treatment experiment done with the Lake Superior Research Institute (LSRI)

in spring of 2018, using spiked water from the DSH (LSRI 2018). However, bacterial regrowth can occur within days after treatment, which re-establishes or exceeds the original densities and shifts community composition (Petersen et al. 2019, Hess-Erga et al. 2019, Waite et al. 2003). Post-treatment nutrient changes can also influence regrowth, as dead organisms can release dissolved organic carbon (DOC) which can stimulate the activity of remaining bacteria (Hess-Erga et al. 2010). A 2019 study by Petersen et al. (2019) concluded that original bacterial populations in ballast water, including indicators and potential pathogens, re-established under appropriate environmental conditions.

Immediate microbial reductions become irrelevant if ballast water is treated at intake and undergoes multi-day holding time before discharge. It is worth noting that all treatment techniques show bacterial regrowth after a given time period, but the scale, timeline, and composition of this regrowth varies with treatment technique, original community composition and environmental factors (Grob and Pollet 2016). Further research is needed to monitor the extent of post-treatment bacterial regrowth, particularly in natural waters outside of the laboratory environment, to better understand the potential impacts of ballast discharge on natural ecosystems (Hess-Erga et al. 2019).

In this project, ultraviolet (UV) radiation, chlorine, and control treatments were applied to ambient and FIB-spiked samples collected from DSH alongside an active and recently docked lake freighter. Culture-based and molecular techniques (qPCR and partial 16S rRNA gene sequencing analysis) were used to quantify indicator bacteria, *E. coli* and *Enterococcus sp,* in order to determine the removal of target species after each treatment. Samples were analyzed immediately after treatment, as well as five days after

treatment to check for bacterial regrowth. The relative abundances of indicator and pathogen-containing genera were also determined to evaluate whether bacterial genera harboring potential pathogens are removed similarly to indicator bacterial groups in each treatment.

The objectives of this project were to (a) compare the removal of indicator bacteria determined by molecular (qPCR) and culture-based methods, (b) determine whether the removal of indicator bacteria is representative of the fate of other potentially pathogenic bacterial genera, and (c) detect the scope of bacterial regrowth within post-treatment ballast water in the Duluth-Superior Harbor.

Methods

Field Sampling

Water samples were collected by boat from alongside an active lake freighter at Canadian National/CD Duluth Dock (46.750261, -92.133328) in August 2019. Along with the four 5-gallon carboys of surface water, general water quality data (i.e., water temperature, pH, dissolved oxygen [DO], water depth, secchi depth) were collected using a YSI EXO2 multi-parameter sonde. Within an hour of collection, the water samples were transported in a chilled cooler to the Lake Superior Research Institute and divided into 1L subsamples for treatment.

Sample Preparation and Treatment

Half of the water samples were spiked with (60 µL of 6-hr log-phase culture per liter) of *Escherichia coli* (ATCC#25922) and *Enterococcus faecium* (ATCC#35667) (Fig. 1).

Spiked harbor water samples were included to ensure that accurate indicator removal could be determined, if indicator bacteria levels were low in the ambient (unaltered) water samples. Both ambient (unspiked) and indicator-spiked samples were treated with chlorine and UV.

For the chlorine treatment samples, one liter of harbor water was used to determine the chlorine demand of this water prior to dosing. Chlorine demand was calculated by subtracting the total residual oxidant concentration from the initial chlorine dose (10mg/L) that was added to the samples and reacted with a DPD Total Chlorine Reagent Powder Pillow. Once the appropriate chlorine dose was calculated, fourteen 1 L samples (7 ambient replicates, 7 spiked replicates) were treated with a bleach solution equal to the chlorine demand of the water + 6 mg/L chlorine. After 30 minutes, the chlorine was neutralized with 45-75µl of 39% w/v sodium thiosulfate, depending on the total residual oxidant (TRO) concentration for each sample. The TRO was checked again following neutralization to ensure that it was effective (<0.021 mg/L).

For the UV treatment, 980ml harbor water subsamples were evenly divided between fourteen quartz test tubes and treated with a UV radiation dose of approximately 100 mJ/cm^2 using a Rayonet™ Merry-Go-Round style reactor Model RPR-100; Southern New England Ultraviolet Company, Branford, Connecticut, USA. After treatment, a portion of the water from each treatment category was used to measure water quality parameters, including temperature, pH, DO, conductivity, percent transmittance, total non-purgeable organic carbon (NPOC) and dissolved organic carbon (DOC) (Table 2). DO was measured using a Hach LDO HQ30d DO meter; conductivity and temperature

with an Oakton Model CON 110 Conductivity/TDS/Temperature Meter; pH with an Orion 3 Star meter and Orion 8157BNUMD pH probe; percent transmittance with a Real UV254 P200 and Perkin Elmer Lambda 35 UV-Vis spectrophotometer; NPOC and DOC with a Shimadzu Model TOC-L Total Organic Carbon Analyzer.

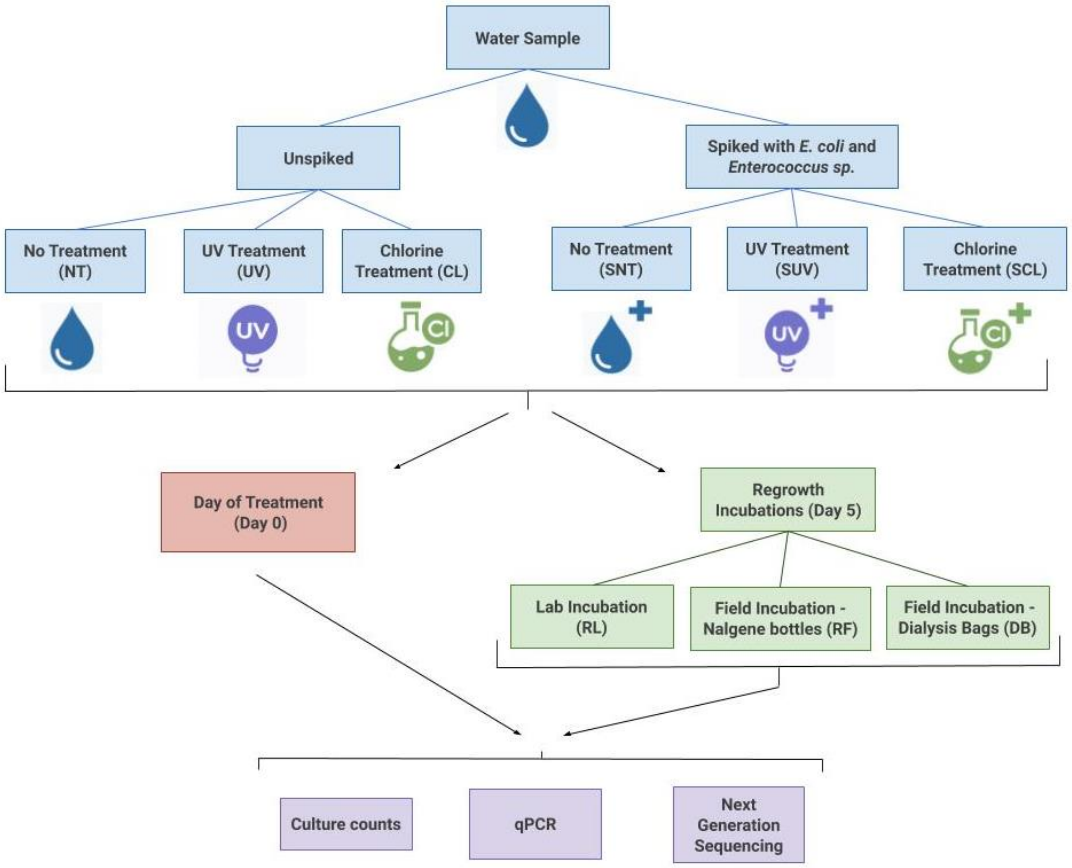


Figure 1. Experimental design of 2019 bench-scale ballast treatment experiment. Water samples were collected from Duluth-Superior Harbor from alongside an active lake freighter.

Post-treatment Sample Processing

Treated water samples were analyzed for abundance of indicator and heterotrophic bacteria, total prokaryotic cells and bacterial community composition. Culturable

indicator and heterotrophic bacteria were enumerated by IDEXX Colilert and Enterolert. Total prokaryotic cells were estimated using DAPI staining and epifluorescence microscopy (Porter and Feig, 1980). For DNA sequencing and qPCR analysis of indicator bacteria, approximately 500 ml of water was filtered through separate Durapore® membrane filters (47 mm diameter, 0.22 µm pore size) to harvest microbial cells. The volumes of water filtered were recorded. These filters were folded (to protect the surface contents), placed in sterile Whirlpak bags, and frozen at -80°C until DNA could be extracted.

Regrowth Incubations

To investigate bacterial regrowth after the UV and chlorine treatments, subsamples of post-treatment water from each treatment category were incubated for five days in laboratory and field settings. An incubation duration of five days was selected as past studies have exhibited major bacterial regrowth after approximately 3-7 days (Hess-Erga et al. 2010, Petersen et al. 2019). For the laboratory incubation, 500 ml replicates were aliquoted into Pyrex bottles and held in an incubator at 20°C, the temperature of the harbor surface water at the time of sampling. The bottles were loosely capped in order to allow for some air to enter. As a dark, closed system, the laboratory incubations served as a proxy for water held in a ballast tank. There were two types of field incubations: a dialysis bag field incubation and a Nalgene bottle field incubation. The dialysis bag incubation served as a proxy for the DSH receiving waters because dialysis bags (12-14kD) allow for water and small ions, but not larger bacterial cells, to move

bidirectionally through the pores in the tubing. For these incubations, two 75 ml post-treatment water from each treatment category were secured in dialysis tubing. To allow for water flow, but protect from physical disturbance, the dialysis bags were enclosed in sterile mesh bags and suspended within a cage of two plastic crates. The container was suspended off Montreal Pier (46.712267, -92.046820) 1 m below the surface of the DSH. The separate Nalgene bottle incubations also exposed the water samples to field light and temperature conditions and served as insurance in case dialysis tubing broke. The Nalgene incubations involved floating 500 ml post-treatment replicates off of a DSH pier 1 m below the surface in transparent Nalgene containers. After 5 days, the post-treatment water samples from the lab and field incubations were processed for the quantification of culturable bacteria and microbial harvesting through filtration for the extraction of genomic DNA.

Culture-based Quantification of Indicator Bacteria

Staff members at the University of Wisconsin-Superior LSRI utilized most probable number (MPN) dilution-culture methods to quantify the following culturable bacteria. Total Coliforms and *Escherichia coli* used IDEXX Colilert and Quanti-tray/2000, and these trays were incubated at 35^{o}C ± 0.5°C for 24-28 hrs prior to data collection. *Enterococcus* spp. used the IDEXX Enterolert and Quanti-tray/2000 and were incubated at 41^{o}C ± 0.5°C for 24-28 hours. Heterotrophic bacteria counts were determined with the IDEXX HPC for Quanti-Tray and Quanti-tray/2000, and incubated at 36^{o}C ± 0.5°C for 48-72 hours.

Total Prokaryotic Direct Cell Counts

Total prokaryotic cells were counted in subsamples of water from each treatment using 4′6-diamidino-2-phenylindole (DAPI) staining and epifluorescence microscopy (Porter and Feig 1980). 10 ml of well-mixed water from each of three replicates (A, B, and C) for the UV, chlorine or regrowth treatments was placed into a liquid scintillation vial, preserved with 37% formaldehyde (1.8% final concentration), and then stored refrigerated (4°C) until cells could be counted (less than 2 weeks). Subsamples (0.2-0.5 ml) from the first three replicates of each treatment (A, B, and C) were stained with DAPI and concentrated onto black polycarbonate filters (Poretics; 25 mm dia., 0.2 µm pore size). For each filter, prokaryotic cells were counted in 10 fields of view at 1,000X total magnification using a Nikon Eclipse 80i epifluorescence microscope with a wavelength of excitation at 365nm and emission at 418nm. These cell counts along with dilution information were used to calculate total prokaryotic cell abundances (cells/ml) in each treatment subsample.

DNA Extraction

DNA was extracted from the Duropore filters containing microbial cells from the different treatments using the Modular DNA Extraction Protocol (from Levar et al. 2015 as updated by C. Sheik and J. Knack 2018). Nuclease free water used in PCR reactions was filtered through a blank filter and DNA was extracted alongside the treatment samples to test for reagent contamination and serve as an extraction blank. DNA

concentrations were quantified using a NanoDrop spectrophotometer and Invitrogen™ Qubit Fluorometer (Thermo Fisher Scientific), and then frozen (-80°C) prior to qPCR reactions and Illumina sequencing.

Quantitative PCR (qPCR)

Bacterial standards were created for qPCR assays using the same *E. coli* and *Enterococcus faecium* stocks used to spike water samples. Bacterial cells were cultured according to ATCC propagation protocols for each strain, pelletized and rinsed 3x with sterile Phosphate Buffered Saline (PBS) and resuspended in 5mL of PBS (USEPA 2014) before DNA was extracted and quantified (as described above for water samples). Serial dilutions (10x) were performed to create DNA standard curves for each target species. The qPCR assay for *E. coli* involved the amplifying and detecting a specific region of the large subunit ribosomal ribonucleic acid (RNA) gene (lsrRNA, 23S rRNA) (USEPA 2014). The concentration of gene copies per µl in the standard *E. coli* genomic DNA preparation were determined from the total DNA concentration and the following formula:

$$\text{Concentration of lsrRNA gene copies/µl} = \frac{Total\ DNA\ conc.(\frac{fg}{\mu l})}{5.06\ fg/genome} \times \frac{7\ lsrRNA\ gene\ copies}{genome}$$

This formula is based on the weight of a single *E. coli* genome (~5.06 fg) and the fact that there are seven lsrRNA gene copies per genome for *E. coli* (rRNA operon database: http://rrndb.cme.msu.edu) (EPA 2014). The same formula was used for *Enterococcus*

faecium except the genome size (2.8fg/genome) and lsrRNA gene copies (6 copies) was changed accordingly.

Primers and probes for the qPCR analyses of *E. coli* and *Enterococcus faecium* are listed in Table 1. The *E. coli* primers (EC23S857) are well established and used in the EPA's "Method C" protocol for quantifying *E. coli* using TaqMan® qPCR (Aw et al. 2019). The primers and probe set for the *Enterococcus faecium* (Cium) qPCR assay have also been verified by past studies (Ryu et al. 2013, Gehring and Santo 2017, Kapoor et al. 2015). Total bacteria were measured using a segment of the 16S rRNA gene (Nadkarni et al. 2002).

Table 1. Primers, probes, and PCR conditions used in the qPCR analyses to determine the presence of indicator and total bacteria. BP refers to the length of the amplicons in number of base pairs and Temp. refers to the annealing temperature for each target gene.

Marker	Target Bacteria	Forward Primer	Reverse Primer	Probe	BP	Temp.	Source
EC23S 857	*Escherichia coli*	5'-GGTAGAGCACTG TTTTGGCA -3'	5'-TGTCTCCCGTGAT AACtTTCTC -3'	FAM-5'-TCATCCCGAC TTACCAACCC G -3-TAMRA	88 bp	56°C	Aw et al. 2019
CIUM	*Enterococcus faecium*	5'-TTCTTTTTCCACC GGAGCTT -3'	5'-AACCATGCGGTT TYGATTG -3'	FAM-5-AGTAACACGT GGGTAACCTG CCCATCAGA -3' TAMRA	141 bp	60°C	Ryu et al. 2013
16S rRNA	Total bacteria	5'-TCCTACGGGAGG CAGCAGT-3′	5' GGA CTA CCA GGG TAT CTA ATC CTG TT 3'	5' CGT ATT ACC GCG GCT GCT GGC AC 3'	466 bp	60°C	Nadkarni et al. 2002

Each qPCR reaction contained 10 µl BioRad iTaq Universal Probes Supermix (Bio-Rad, CA, USA), 1 µl of each 10 µM primer, 0.4 µl of probe, 2.6 µl nuclease-free water and 5 µl of DNA template for a final reaction volume of 20 µl per well. A standard curve was run alongside the samples. Extraction blanks and non-template PCR blanks

(nuclease free water) were also included on each qPCR microplate to serve as negative controls.

The StepOnePlus thermocycler (ThermoFisher, MA, USA) program for *E. coli* was also based on EPA's most recent draft of "Method" (Aw et al. 2019). Thermal cycling protocols were 50°C for 2 min, 95°C for 10 min, 40 cycles of 95°C for 15 sec followed by 56°C for 1 min. The amplification conditions were the same for *Enterococcus faecium* and total bacteria (16S rRNA), except for the annealing temperature which was set to 60 °C for 1 min.

Data Analysis of Abundance

Bacterial abundance results from molecular and culturable methods were used to calculate percent removal for each ballast treatment technique. Percent removal was calculated using the following equation (where is A_b is abundance before treatment and A_a is abundance after treatment):

$$\text{Percent removal} = \frac{(A_b - A_a)}{A_b} \times 100$$

Removal and regrowth were analyzed using t-tests, analysis of variance (ANOVA) and Tukey's post hoc comparison of means (Tukey's HSD) statistical methods to determine whether observed differences were significant. The threshold of significance throughout these analyses was $p<0.05$.

16S rRNA Sequencing and Bioinformatics

To gain information about the bacterial community composition of each sample, extracted DNA was sent to the University of Minnesota Genomics Center (Minneapolis, MN, USA) for sequencing of the V4 region of 16S rRNA gene using an Illumina MiSeq Version 3 Chemistry DNA sequencer (Gohl et al. 2016, Hamilton et al. 2013, Schloss and Handelsman 2005). The DNA was amplified using qPCR to control cycles and limit PCR chimera formation using a primer pair (Meta_V4_515F; 5'GTGCCAGCMGCCGCGGTAA and Meta_V4_806R; 5'GGACTACHVGGGTWTCTAAT). The amplified product was cleaned and the prepared library were sequenced.

The 16S rRNA sequences were trimmed and aligned against the SILVA SSU rRNA database (v. 1.32) using mothur (v. 1.43.0) and sequences were clustered into operational taxonomic units (OTUs) at a 97% similarity level using OptiClust. Taxonomy was assigned at the genus level using the same SILVA reference database. Sequence data was imported into R (v. 3.5.0) and analyzed using phyloseq and ALDEx2 packages (McMurdie and Holmes 2013). Rarefaction curves were calculated to quantify the completeness of sampling. Comparisons of bacterial constituents between treatments was determined by examining the numbers and types of OTUs.

To analyze diversity, we scaled the OTU counts by bootstrapping counts of random samples drawn at the lowest read depth (30,323 reads) 100 times. These OTU counts were then used with scaled Shannon, Simpson and Chao1 diversity indices to calculate evenness and richness. Differences in community composition were visualized with a

phyloseq-based principal coordinate analysis (PCoA) using the Jaccard distance index (McMurdie and Holmes 2013). A permutational multivariate analysis of variance (PERMANOVA) through the *Adonis* function in R was used to statistically test the community differences between treatment groupings by comparing the centroid and dispersion of each cluster (Oksanen et al. 2011). The relative abundance of 136 bacterial pathogen-containing genera were compared in bacterial communities found in different UV, chlorine, and regrowth treatments.

Results

Water Quality Data

The temperature, pH, dissolved oxygen and secchi depth of the water taken from the Duluth-Superior Harbor for the treatment and regrowth experiments were comparable to values previously reported for this estuary (GLMRI 2012). The water temperature at the sampling location (46.750261, -92.133328) was 20.7°C, DO was 8.55 mg/L, secchi depth was 0.69 meters and the water depth was 5.1 meters. The post treatment water quality parameters (Day 0) indicated that chlorinated samples had the highest conductivity, DOC and transmittance (Table 2).

Table 2. Water quality parameters measured post-treatment (Day 0).

Treatment Category	Temperature (°C)	pH	Dissolved Oxygen (mg/L)	Conductivity (µS/cm)	NPOC (mg/L)	DOC (mg/L)	Transmittance % at 254nm
No-Treatment (NT)	17.1	7.85	9.2	168	9.88	9.95	34.8
Chlorine (CL)	18.9	8.33	9.0	248	9.88	10.2	38.5
Ultraviolet (UV)	21.2	7.88	9.1	160	9.90	9.85	35.5

Bacterial Abundance Before and After Treatment

The concentration of culturable heterotrophic bacteria in the harbor water was 5.1 × 10^4 MPN/100mL (±5,848 SEM) using the IDEXX HPC for Quanti-Tray system. After treatment with UV and chlorination, culture-based results revealed greater than a 99% reduction in all indicator bacteria (*E. coli*, *Enterococcus sp.*, total coliforms and heterotrophic bacteria) in all samples (Fig. 2). Using qPCR, quantification of total bacteria in the untreated harbor water (culturable and nonculturable) was 5.3 × 10^8 16S rRNA copies per 100 mL. The percent removal of *E. coli* gene copies in unspiked samples was 67.5% and 96.8%, respectively, for the UV and chlorine treatments, and 91.8% and 99.9% in spiked samples (Table 3). Percent removal, as quantified by qPCR, was lower for the unspiked samples than for the more concentrated spiked samples. The samples treated with UV radiation, showed the largest difference between culture-based and qPCR quantification. Unpaired t-tests and Tukey's comparison of means (Tukey's HSD) were used to compare changes in bacterial abundance in untreated and treated samples for both the culturable and qPCR detection methods. These tests confirmed the removal of bacteria after treatment for both treatment techniques ($p<0.0001$) in both

spiked and unspiked samples. Total prokaryotic cell abundance of untreated harbor water using microscopy was 9.58×10^{8} cells per 100ml; however, results from a one-way ANOVA did not show a significant difference in the number of intact cells between treatment categories (F-stat=1.269, df=5, p=0.339).

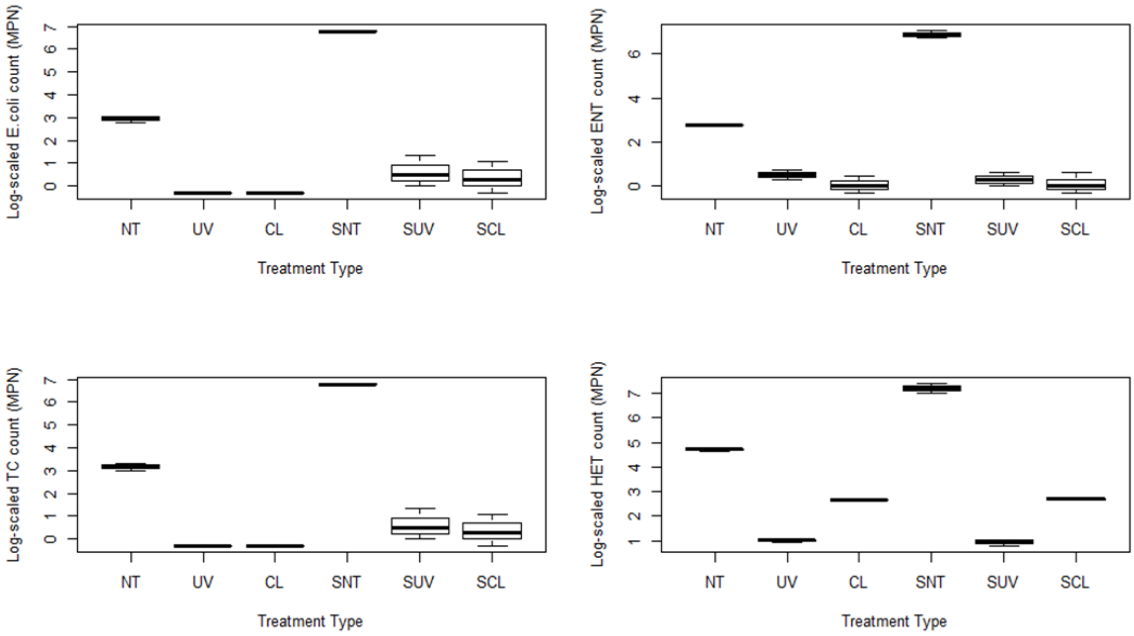


Figure 2. Log-scaled boxplots of culturable counts (MPN) of treatment-day samples for indicator and heterotrophic bacteria found in various treatments: *E. coli* (top left), *Enterococcus* sp. (top right), total coliform (bottom left) and heterotrophic bacteria (bottom right). Treatment categories included: no treatment (NT), UV light (UV), chlorine (CL), spiked no treatment (SNT), spiked-UV (SUV), and spiked-chlorine (SCL).

Table 3. Percent removal (%) of indicator and total bacteria in different treatments, quantified through qPCR analysis and microscopy. Percent removal was calculated using the average abundance of triplicate samples from each treatment category.

Method	Target	Unspiked UV	Spiked UV	Unspiked Chlorine	Spiked Chlorine
qPCR	*Escherichia coli*	67.5	91.8	96.8	100.0
	Enterococcus faecium	77.2	95.0	99.2	100.0
	16S rRNA (total bacteria)	99.1	99.6	99.8	99.9
DAPI	Prokaryotic cells	31.1	19.4	36.4	7.9

Regrowth Incubations

To explore post-treatment bacterial regrowth, the bacterial populations of the laboratory and field incubations were quantified after five days. The total bacteria in spiked samples, as measured by 16S rRNA abundance, increased 4-6 orders of magnitude after the chlorine treatment, exceeding the original pre-treatment concentration (Fig. 3i). Spiked samples treated with UV light also showed an increase in total bacterial 16S rRNA gene copies by 3-4 orders of magnitude (Fig. 3h). There was no regrowth of *E. coli* cells after 5 days in the UV treatment (Fig. 3b), however, our results showed notable regrowth of *E. coli* (Fig. 3c) and total bacteria (3.i) in samples treated with chlorine. Specifically, *E. coli* abundance increased 2-4 orders of magnitude from the post-treatment concentration values after the chlorine treatment (Fig. 3c). The abundances of *Enterococcus sp.* remained low after 5 days for both the UV and chlorine treatments indicating no cell regrowth (Fig 2e and f). *E. coli* (EC) and *Enterococcus sp.* (ENT) abundances in the untreated control (NT) decreased by 2-4 orders of magnitude after 5

days in each incubation environment, but the abundances of total bacteria (i.e., bacterial 16S rRNA gene copies) were similar in the different incubation environments and as high as immediately after treatment (Fig. 3a, d, g).

Considering differences in cell abundance between the different incubation environments, there were no differences in the abundance of total bacteria (Fig. 3h; ANOVA, F-stat=0.269, df=2, p=0.77) or *E. coli* (Fig, 3b; ANOVA, F-stat=3.088, df=2, p=0.063). The only significant difference in regrowth environment was the abundance of *Enterococcus* between the laboratory regrowth and the dialysis bag incubations (Fig. 3e; Tukey's HSD, p=0.027).

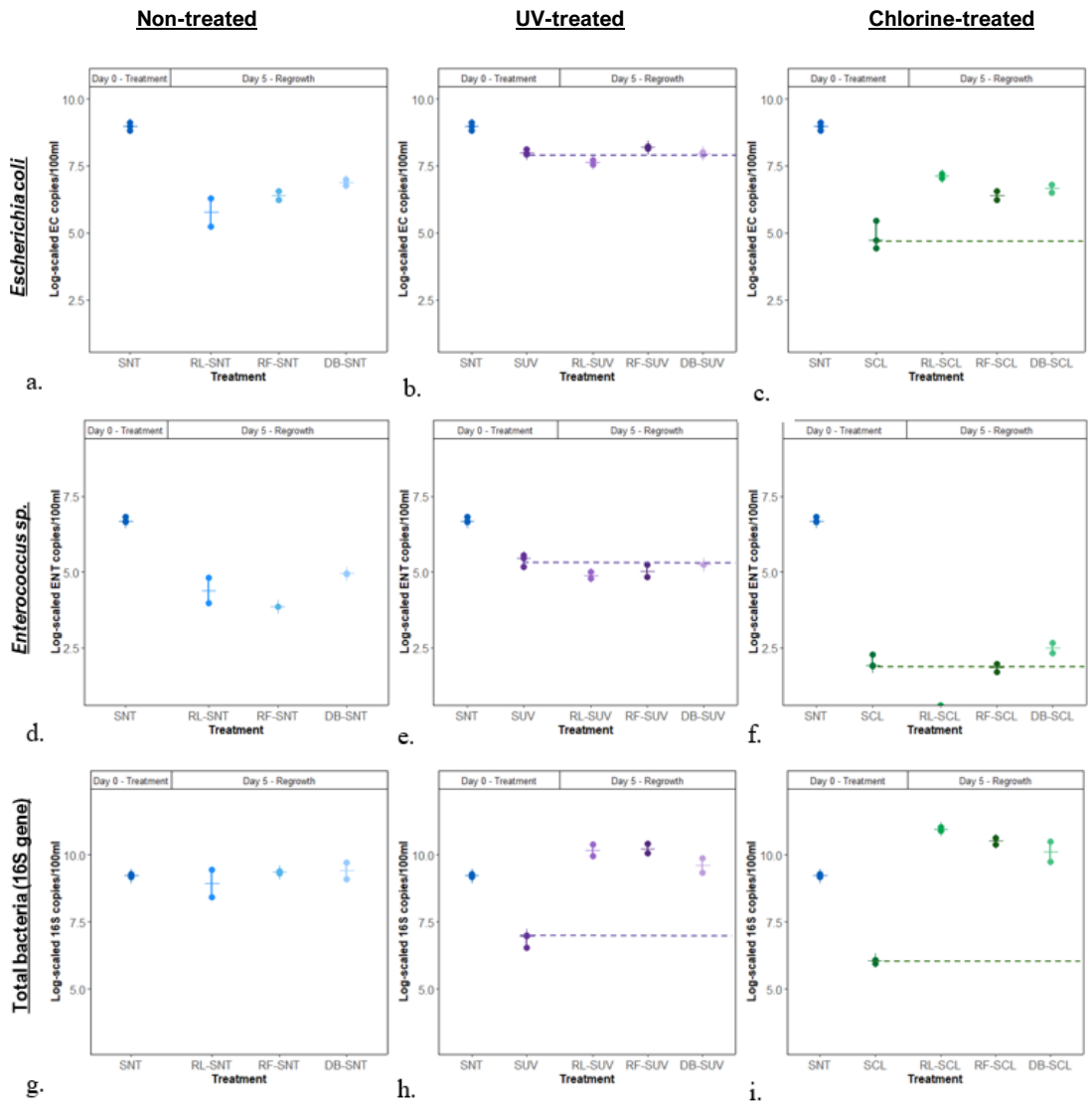


Figure 3. Abundance of target bacteria in spiked samples determined by qPCR analysis. Log-scaled abundance of *E. coli* (target gene copies/100 ml) (Figures 2a-c), log-scaled abundance of *Enterococcus sp.* (target gene copies/100ml) (Figures 2d-f), (c) log-scaled abundance of total bacteria (16S rRNA gene copies/100ml) (Figures 2g-i). Dashed line represents the median abundance immediately after treatment (Day 0). Day 5 regrowth samples include three incubation environments: lab incubation (RL), Nalgene field incubations (RF), dialysis bag field incubations (DB). Treatment categories include spiked non-treated samples (SNT), spiked and UV-treated (SUV), and spiked and chlorinated (SCL).

Bacterial Diversity and Community Composition

A mean of 2776 ± 390.5 OTUs were identified in the untreated water taken from the Duluth-Superior Harbor, with the classification to 110.6±3.8 orders and 20.3±0.6 phyla. Among all samples, 1.73% of sequence reads could not be classified to an order. The bacterial community composition of the untreated water sample was diversely comprised of members of the orders of Betaprotobacteriales (12%) and SAR11 group of marine alpha-proteobacteria (5.1 %) in the phylum Proteobacteria, Frankiales (11%) and Microtrichales (9.2%), in the phylum Actinobacteria, and Chitinophagales (5.3%), Sphingobacteriales (4.4%) and Flavobacteriales (3.3%) in the phylum Bacteroidetes (Fig. 4). All other orders accounted for a mean of < 49.7 % of sequence reads.

The compositions of bacterial communities in untreated and UV-treated samples were similar at the end the initial treatment day (Day-0), while the communities in the chlorinated samples were noticeably different (Figs. 3 & 4). Both species richness and evenness of bacterial communities decreased after chlorine treatment ($P < 0.001$; Table 5 and Fig. S3). An average of 685.3 ± 62.5 OTUs were identified among the chlorinated samples, along with lower Shannon indices (a measure of richness and evenness) and Simpson diversity indices decreased ($p<0.001$). This reduction in species diversity did not occur after the UV treatment (Table 5). After chlorination, the predominant remaining phyla were Planctomycetes, Cyanobacteria, and Proteobacteria (in descending order by relative abundance). The most prevalent phyla of the untreated and UV-treated samples were Proteobacteria, Actinobacteria, and Bacteroidetes.

Ordination of Jaccard distances by principal coordinate analysis (PCoA) revealed clustering by time after treatment (Day 0 vs. Day 5) and treatment type in both unspiked and spiked samples (Fig. 5). The post-treatment communities from chlorinated samples were initially very different from those of the UV and non-treated samples. However, after five days, there was a distinct shift in the bacterial community composition (Fig. 5) in both unspiked and spiked samples. Specifically, after 5 days the phylum of Proteobacteria increased in the treated samples, particularly in those treated with chlorine (Fig. 4). The bacterial community compositions between the three regrowth environments were not significantly different (PERMANOVA, $p=0.134$)

Table 4. Diversity indices based on bacterial community partial 16S rDNA sequences for water in different treatments (NT, UV CL, SNT, SUV, and SCL) and regrowth incubation environments (RL, RF, and DB). Average and standard deviation of each metric across the replicates of each sample type.

Sample Type	Chao1	SD	Shannon	SD	Simpson	SD
NT	1391.3 ±	128.1	5.09 ±	0.06	0.99 ±	0.00
UV	1317.6 ±	173.5	5.07 ±	0.12	0.98 ±	0.00
CL	622.1 ±	158.7	3.41 ±	0.48	0.89 ±	0.07
SNT	1242.8 ±	69.1	5.03 ±	0.00	0.98 ±	0.00
SUV	1331.0 ±	101.0	4.89 ±	0.08	0.98 ±	0.00
SCL	813.7 ±	74.4	3.77 ±	0.16	0.92 ±	0.03
RL-NT	1448.3 ±	83.3	5.09 ±	0.08	0.98 ±	0.00
RL-UV	710.5 ±	98.3	3.18 ±	0.55	0.89 ±	0.05
RL-CL	305.1 ±	8.3	2.49 ±	0.20	0.84 ±	0.04
RL-SNT	1556.0 ±	163.5	4.96 ±	0.05	0.98 ±	0.00
RL-SUV	543.4 ±	0.0	2.99 ±	0.00	0.91 ±	0.00
DB-SNT	1329.1 ±	0.0	4.33 ±	0.00	0.96 ±	0.00
DB-SUV	808.5 ±	0.0	3.46 ±	0.00	0.93 ±	0.00
DB-SCL	1013.9 ±	0.0	2.87 ±	0.00	0.90 ±	0.00
RF-NT	1236.0 ±	146.0	4.97 ±	0.12	0.98 ±	0.00
RF-UV	812.6 ±	151.9	3.83 ±	0.18	0.94 ±	0.01
RF-CL	431.3 ±	253.1	2.73 ±	1.06	0.86 ±	0.11
RF-SNT	1240.7 ±	91.0	4.68 ±	0.16	0.97 ±	0.00
RF-SUV	539.2 ±	25.3	2.55 ±	0.81	0.83 ±	0.11
RF-SCL	379.7 ±	197.5	2.52 ±	0.93	0.85 ±	0.09

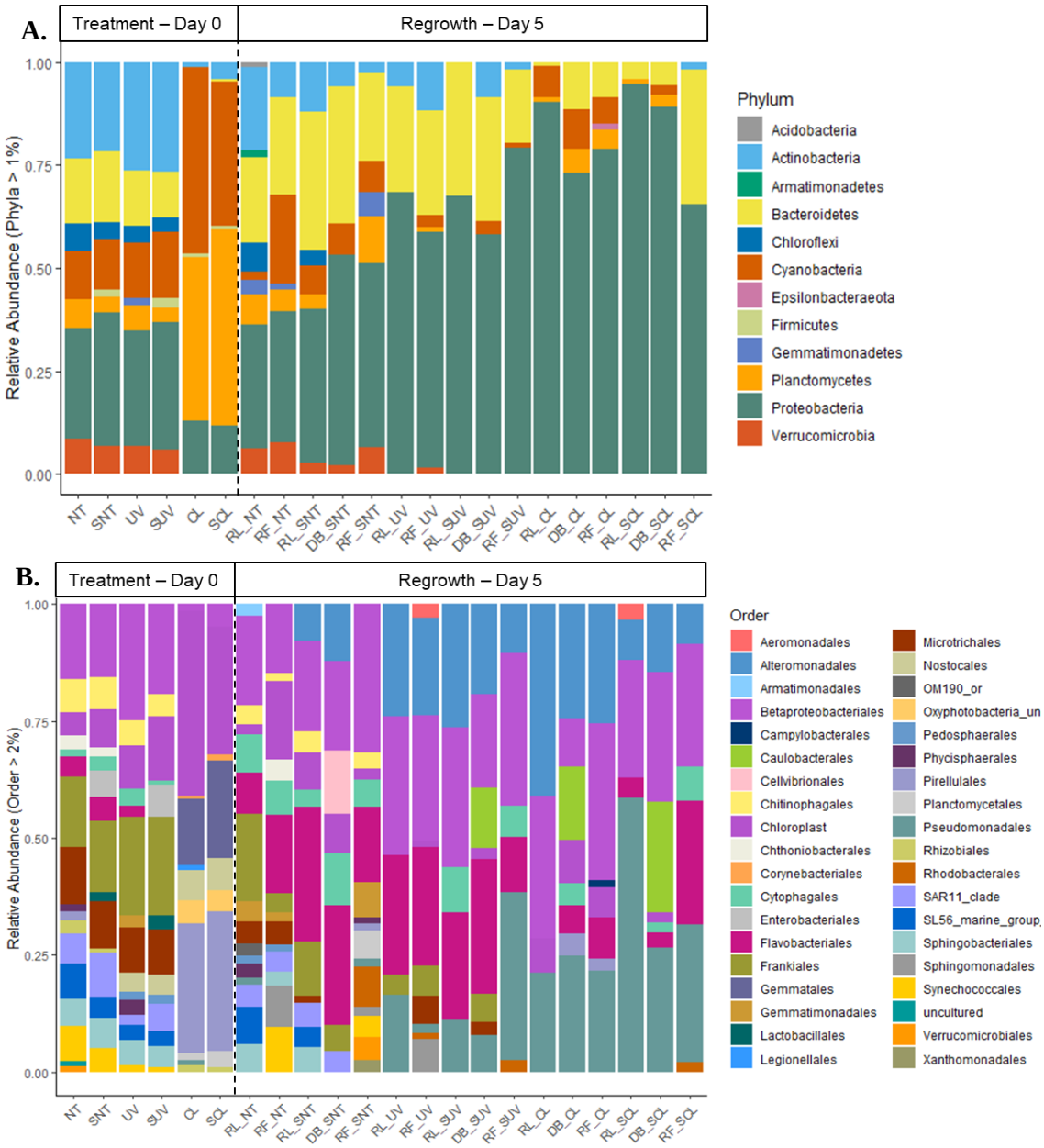


Figure 4. Relative abundance (%) of partial 16S rRNA sequences from bacterial communities at the (a.) Phylum level and (b.) Order level. Same day treatment samples (Day 0) on the left include unspiked non-treated (NT), UV-treated (UV), and chlorine-treated (CL) samples, as well as their spiked counterparts (SNT, SUV, SCL). The regrowth samples (Day 5) on the right include samples from three different incubations: a lab incubation (RL), a field incubation in dialysis bags (DB), and a field incubation in Nalgene bottles (RF).

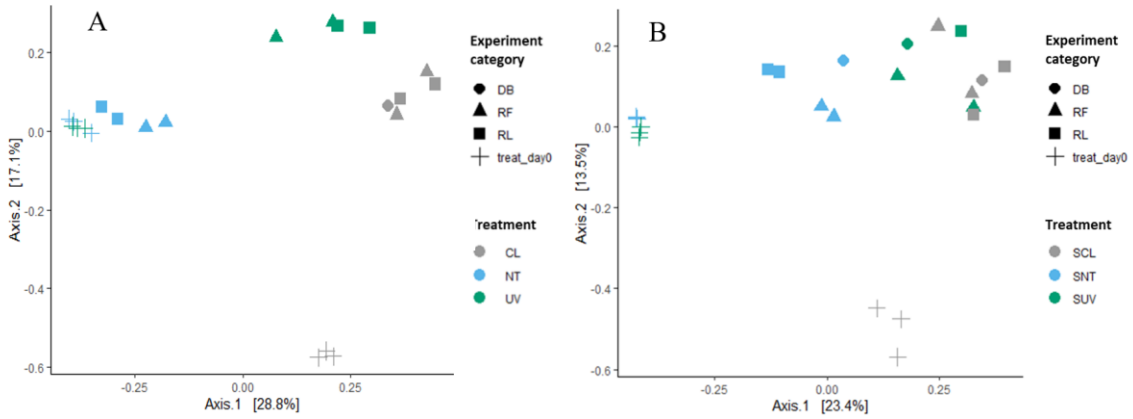


Figure 5. Principle Coordinate Analysis (PCoA) of partial 16S rRNA sequences from bacterial communities in (a.) unspiked and (b.) spiked dock water samples using the Jaccard distance index. The Day 0 samples are represented by + symbols while the Day 5 regrowth samples are represented by squares (lab incubation – RL), triangles (nalgene field incubations – RF) and circles (dialysis bag field incubations – DB). Treatment categories include non-treated samples (NT, SNT), chlorinated (CL, SCL), and UV-treated (UV, SUV).

Pathogen Analysis

At the end of the treatment day (Day 0) samples, members of pathogen-containing genera (PCGs) accounted for about 1% of the bacterial communities in all samples (unspiked, non-treated and treated samples (Fig 6). For spiked non-treated and spiked UV-treated samples, about 8% of the total sequences were contributed by PCGs (mostly the *Enterococcus sp.* and *E. coli* used to spike the samples). For samples treated with chlorine (both spiked and unspiked) PCGs accounted for 5-6% of all sequences. For chlorinated samples, sequences from the *Legionella* and *Mycobacterium* bacterial genera were the dominant sequences from PCGs.

After five days, the relative abundances and varieties of PCGs within the bacterial communities had changed dramatically in all treatment groups compared to the initial

treatment day (Fig 6). The most dramatic shift in proportions occurred in the spiked chlorinated samples where the relative abundance of different PCGs increased from 5% to 60% after the 5-day laboratory incubation. The smallest shifts in the relative abundances of PCGs occurred in the non-treated samples. The primary PCGs in the regrowth samples were *Flavobacterium*, *Acinetobacter*, and *Pseudomonas*. The indicator bacteria in the spiked untreated samples disappeared almost completely (fell below 0.1% relative abundance) in the regrowth samples. These indicator bacteria did not regrow after treatment, while bacteria from many other and often different genera did.

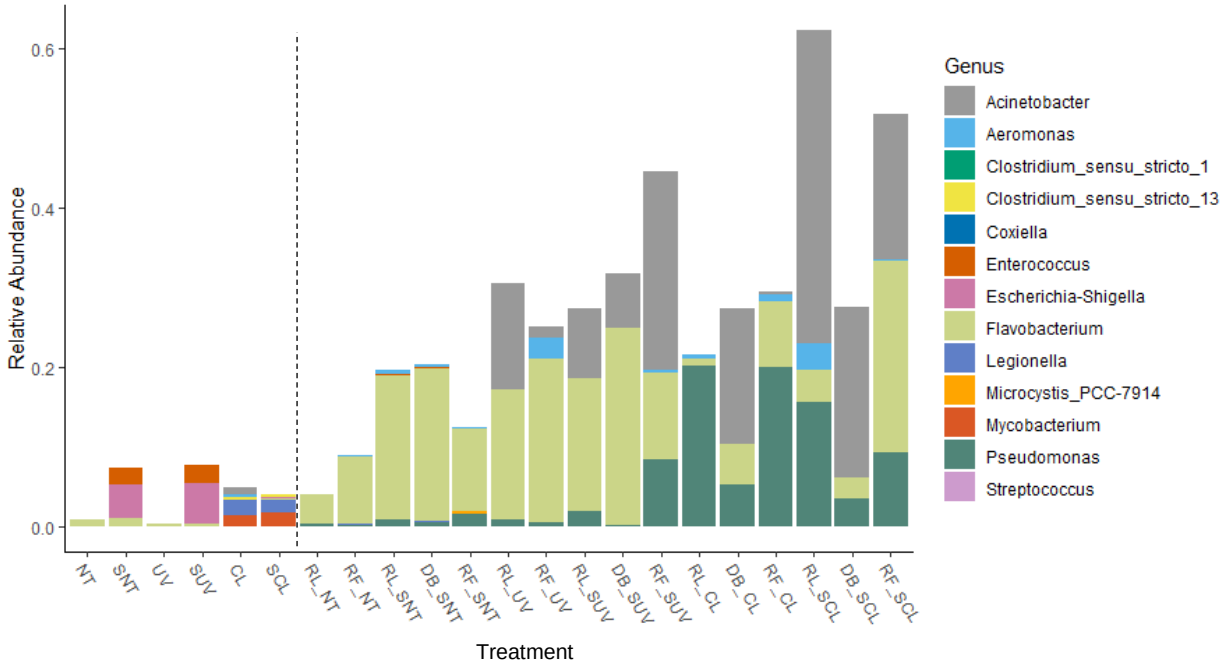


Figure 6. Relative abundance of pathogen containing genera determined from 16S rRNA sequences in the different treatments at the conclusion of the removal experiment (lab – left of vertical dashed line), and in different conditions within various regrowth incubation environments at the end of the five-day regrowth experiment (lab and field results – right of vertical dashed line). Dashed line separates samples from the different treatments at the end of the lab removal experiment (Day 0), on the left, from the samples from these water treatments in the three regrowth environments, on the right (Day 5). The regrowth experiment included water samples from three different types of incubation environments: a lab incubation (RL), a field incubation in dialysis bags (DB), and a field incubation in Nalgene bottles (RF).

Discussion

Overall, culture-based quantification, qPCR, and DNA sequencing demonstrated that both UV-radiation and chlorination effectively reduced indicator bacteria in the sample water. The differences in the percent removal results between each method of quantification (culture-based vs. molecular) likely reflect the differences between the variables quantified by each technique (live, culturable bacteria vs. amplifiable genetic information). The removal of *E. coli* and *Enterococcus* between the two techniques was more similar (and effective) for spiked samples than for the unspiked samples, which showed lower removal through qPCR (Table 3). The reduced efficacy may be related to the low initial abundance of indicator bacteria in the unspiked harbor water. While culturable methods may be better for instances where quantifying live indicator bacteria with easy and established protocols is preferred, molecular methods are able to account for VBNC cells and can provide information on wider range of microbial organisms and potential pathogens. DAPI staining is helpful for visualizing bacterial cells, however, it did not reflect the post-treatment removal indicated by the other techniques. One possible reason for this is that DAPI has been found to bind to inactive bacterial cells without nucleotides known as "ghost cells" (Saby et al. 1997, Zweifel & Hagstrom 1995). Staining of "ghost cells" in treated samples would have resulted in an overestimate of prokaryotic cells and a reduction in the calculated removal rate.

Despite the removal of over 99% of total bacteria seen in culturable and molecular methods (Fig. 2, Table 3), the composition of the bacterial communities in the non-treated and UV-treated samples was similar (Fig. 4). This result indicates

proportional disinfection across bacterial taxa when UV light is used as a disinfectant. Conversely, bacterial communities in the chlorinated samples were different from the non-treated samples, indicating that some bacterial taxa were more resistant to chlorine disinfection than other taxa. The genus *Mycobacterium* was one of the most prevalent PCGs in the chlorine-treated samples (Fig. 6) and this genus is known for its chlorine resistance. Strains of *Mycobacterium avium* have been found to be at least 500 times more resistant to chlorine than *Escherichia coli* (Taylor et al. 2000, Falkiham 2003). Planctomycetes and Cyanobacteria, the phyla with the largest relative abundance after chlorination (Fig. 4), also include representatives that have exhibited chlorine resistance (Pang et al. 2006).

UV light deactivates cells through the accumulation of DNA damage, while chlorination breaks the chemical bonds in the molecules of cellular components (e.g. cell walls and membranes) (Byappanahalli 2012 et al.). Different bacteria have different cellular compositions that can influence their sensitivity to chlorine. For example, the cell walls of *Mycobacterium* contain a waxy substance which makes it more hydrophobic, a physical barrier which increases its chlorine resistance (Luo et al. 2021). These differences in cellular composition contribute to the disproportional effects of chlorination on different groups of bacteria.

Bacterial regrowth can occur when bacteria that remain viable after treatment reproduce, when cells are resuscitated from a viable but nonculturable state, or when DNA damage is repaired (Wang et al. 2021). Presumably, the cellular debris of the killed bacteria may provide nutrient sources to stimulate the regrowth of other cells in the

communities after the UV and chlorine treatments in these experiments (Fig. 3), particularly members of the *Psuedomonas* and *Actinobacteria* genera (Fig. 6). In addition, enzymes released from the cells of dead microbes may partially degrade organic matter in water, thereby making it more bioavailable to the remaining bacteria (Hess-Erga et al. 2019, Hess-Erga et al. 2010). We did not see major differences in DOC between treated and non-treated samples (Table 2); however other studies have shown dissolved organic carbon (DOC) spikes immediately after treatment which eventually stabilizes as the dissolved organic matters is used by opportunistic bacteria (Hess-Erga et al. 2010). Additionally, UV-radiation has been shown to photodegrade existing DOC increase its biolability and stimulate bacterial growth (Hendrika et al 2003).

Some species in the *Pseudomonas* and *Acinetobacter* genera, the two PCGs with the largest relative abundance in the chlorinated samples, are chlorine resistant (Luo et al. 2021). However, unlike the chlorine-resistant *Mycobacterium*, cells within the *Pseudomonas* and *Acinetobacter* genera are likely fast-growing r-strategists (opportunistic organisms that thrive in low competition environments) (Juteau et al. 1999, Andrews & Harris, 1986, Hess-Erga 2010). Organisms that thrive in more crowded environments with strong resource competition (K-strategists) may have been more abundant in the untreated control samples.

The most specific taxonomic classification possible for the partial 16S rRNA sequence analysis used here was genus-level due to the high 16S rRNA gene similarity between closely related species. Therefore, the pathogenicity of the species and strains that occurred within the PCGs detected in these experiments could not be determined.

Both the *Pseudomonas* and *Acinetobacter* bacterial genera, however, contain opportunistic human pathogens and the *Flavobacterium* genus accounts for 13% of total bacterial fish pathogens (de Bentzmann & Plésiat 2011, Verma & Rathore 2015). Although these experiments demonstrated that some members of pathogen-containing genera can escape disinfection with UV light or chlorine, and even regrow and flourish after treatment, the public health and ecological risks of pathogen exposure via ballast discharge require further research.

The dialysis bag field incubations provided insights about what bacterial community changes may occur after treated ballast water is discharged into freshwater harbors or estuaries. However, there are limitations to this proxy. First, the dialysis bag field incubations did not account for bacterial predation by other microorganisms. Second, while harbor water and ions outside the dialysis tubing were able to move into and out the dialysis bag, the bag volumes were small and diffusion rates were probably much slower than if there were no flow barrier. Despite these possible shortcomings, similar changes in bacterial communities were observed in the three incubation approaches despite differing environmental conditions (i.e., light and temperature regimes, physical disturbance, etc.). This similarity indicated that the two types of disinfection treatments and the duration of holding after treatment had greater influences on bacterial community compositions in this study.

It is important to note differences in the application of treatment techniques that may be used in a shipboard setting. Despite the diversity in ballast management systems, chlorination is common and usually only applied once while UV-radiation treatments

often occur during both intake and discharge of ballast water. This double treatment provides any bacteria cells that may regrow with a second exposure immediately before entering the receiving waters (Petersen et al. 2019). Thus, the results reported here support the conclusion that the timing of disinfection plays a crucial role in bacterial abundance of post-treated ballast water. To maximize the impact of disinfection, treatment should occur just prior to discharge.

In conclusion, while current culture-based methods to detect indicator bacteria reduce the cost and complexity to monitor some potentially harmful bacteria in ballast treatment systems, caution should be used when making decisions based on these indicators or detection methods because the results reported here show that the fates of indicator bacteria do not necessarily represent those of other bacterial cells in some pathogen-containing genera. Both UV-treatment and chlorination resulted in >99% removal in culturable indicator bacteria immediately after treatment. However, each indicator responded differently to regrowth conditions with no regrowth in *Enterococcus*, moderate regrowth in *E. coli* for chlorine treated samples and major regrowth in total bacteria for both treatments. Five days after treatment, there was a shift in the overall bacterial community composition. This shift included regrowth of cells in genera that harbor pathogens (particularly the *Acinetobacter*, *Flavobacterium*, *Pseudomonas* genera). These shifts were seen in all regrowth samples and the type of regrowth incubation did not have a significant impact on the bacterial communities that developed within 5 days. These results can be used to inform ballast water bacterial monitoring techniques and

ballast management decision makers as they assess the risks and treatment options regarding potentially harmful microbes.

Chapter 2: Determination of Fecal Contamination in the Skunk Creek and Agate Bay Watersheds

Summary

Fecal contamination of Minnesota's recreational beaches and waterways continues to be a widespread and pervasive problem. Skunk Creek, Burlington Bay and Agate Bay Beach in Two Harbors, MN are listed as "impaired" for the fecal indicator *Escherichia coli*, and recreational advisories are an issue of concern for both residents and tourists alike. This project aimed to differentiate between sources of fecal contamination within the Skunk Creek and Agate Bay watersheds in Two Harbors using culture-independent microbial source tracking methods, and to explore the relationships between elevated levels of *E. coli* and the ancillary water quality parameters along a watershed gradient. Water samples were collected from eight sites along the impaired waterways during base flow and storm events. The levels of *E. coli* were measured along with physicochemical water quality parameters. Potential fecal sources were determined using quantitative polymerase chain reaction (qPCR) analysis with human (HB and Lachno3) and avian biomarkers (GFD) and 16S rRNA sequencing in conjunction with the SourceTracker program. Our results identify hotspots and potential sources of fecal contamination, showing greater levels of human biomarkers in the Agate Bay watershed and a correlation between *E. coli* levels and turbidity in Skunk Creek. Additionally, levels of *E. coli* were correlated with stormwater events and turbidity, while human fecal sources appear to be site-specific and independent from storm events. These findings are beneficial in developing mitigation and management strategies to *E. coli* impairments and applicable to other Minnesota streams experiencing similar stressors.

Introduction

Fecal contamination is one of the leading causes of impairments for streams, rivers and estuaries across the United States (USEPA 2009). Each year, fecal contamination of recreational waterways results in approximately 90 million illnesses and economic costs of $2.2- $3.7 billion nationwide (DeFlorio-Barker et al. 2018). Those values do not account for tourism lost due to recreational beach closures and advisories. Traditional fecal indicator bacteria (FIB), like *E. coli* and Enterococcus, can denote the presence of fecal contamination, but cannot distinguish between sources (Feng et al. 2018, Grabow 1986).

Common sources of fecal material into environmental waterways include agricultural runoff, wildlife deposits, inadequate wastewater treatment, aging sewer infrastructure and faulty septic systems (Johnson et al. 2004, Okabe et al. 2007) Identifying the source(s) of contamination is important as the fecal source can influence the scale of associated public health risks. For example, high concentrations of indicator bacteria from human sewage generally pose a greater potential health risk than non-human sources due to human-specific pathogens associated with sewage-impacted water (Soller et al. 2014). Therefore, identification of human fecal inputs is of particular interest regarding mitigation of public health risks.

In addition to fecal inputs, it is important to note that *E. coli* can become naturalized in stream sediment, nearby soils, and periphyton communities (Ishii et al. 2006, Ksoll et al. 2007). These naturalized populations have been documented in streams

and nearshore areas in the Lake Superior basin. Since naturalized *E. coli* are able to survive and reproduce in the environment, the presence of naturalized *E. coli* does not necessarily coincide with fecal contamination and its associated health risks, leading to erroneous advisories.

Appropriate action cannot be taken to remediate fecal impairments if the origin of the problem is unknown. Microbial source tracking (MST) is a tool to trace the origin of fecal inputs using microbiological, genotypic, phenotypic, and chemical methods (Scott et al. 2002). The most common MST methods currently involve using host-specific molecular markers in qPCR which target the 16S rRNA gene of bacteria associated with the host's gut microbiome (Feng et al. 2018, Green et al. 2012, Kildare et al. 2007). More recently, Next Generation Sequencing (NGS) has become a popular tool for MST (Brown et al. 2017). The SourceTracker program is a Bayesian approach to MST which utilizes NGS gene libraries to determine the proportional contributions of bacteria from a variety of sources to a given sink (McGhee et al. 2020).

While large tributaries receive the most monitoring and regulatory attention due to their large hydraulic loads, small Great Lakes tributaries can have a disproportionately large impact for their size (Mooney et al. 2020). A study on the tributaries of Lake Michigan showed that small tributaries contributed outsized loads of bioavailable nutrients to the Great Lakes system (Mooney et al. 2020). In addition, the nutrients, sediment and fecal inputs that small tributaries convey tend to be retained along the shore because small tributaries lack the momentum of larger waterways (Rueda et al. 2007). These stream inputs may fuel coastal algal blooms and elevate public health risks on

recreational beaches when fecal contamination is present, which emphasizes the need to monitor small streams. Additionally, many MST studies have been performed in urban areas (Templar et al. 2016, Feng et al. 2018, Kapoor et al. 2015), however, fewer have been done in non-urban streams and watersheds.

The primary objectives of this study were to (a) understand spatial and temporal patterns of *E. coli* in a small, non-urban stream watershed, and (b) differentiate between sources of fecal contamination within the target watershed using library independent (molecular biomarkers) and library dependent (SourceTracker) microbial source tracking methods. Alongside these primary objectives, this study also explored the relationships between elevated *E. coli* abundance and physiochemical water quality parameters along a watershed gradient. In addition to helping identify sources and solutions for the watershed in this study, these findings can provide insight and a template for studies in other small, Great Lakes watersheds with fecal impairments.

Methods

Study Site

Two Harbors, Minnesota is a small town (population 3,541) on the north shore of Lake Superior (US Census 2018) (Fig. 7). It includes two recreational beaches, Burlington Beach and Agate Bay Beach, both classified by the Minnesota Pollution Control Agency (MPCA) as "impaired" for *E. coli* (MPCA 2018). Skunk Creek, the 2.7-mile creek that flows through Two Harbors into Burlington Bay, is listed as impaired for both *E. coli* and turbidity (MPCA 2018). The Skunk Creek watershed includes 1,319 acres of forest

(43%), development (39%), pasture-land (9%), and wetland (2%). The relatively high frequency of beach advisories in Two Harbors is a point of concern for residents and summer visitors to the North Shore. Due to *E. coli* and turbidity impairments, reports by the MPCA have recommended regular monitoring of streams and sewer outfalls upstream of the impaired locations to help identify the sources of contamination (MPCA 2018).

Sample Collection

Water samples were collected in duplicate from eight sites on five sampling dates (3 base flow periods, 2 storm-events) during summer 2019. Storm-event samples were taken within 24 hours of local rainfall greater than one inch (Fig. 7). Water temperature, dissolved oxygen (DO), conductivity and turbidity tube measurements were recorded in the field at each sampling location.

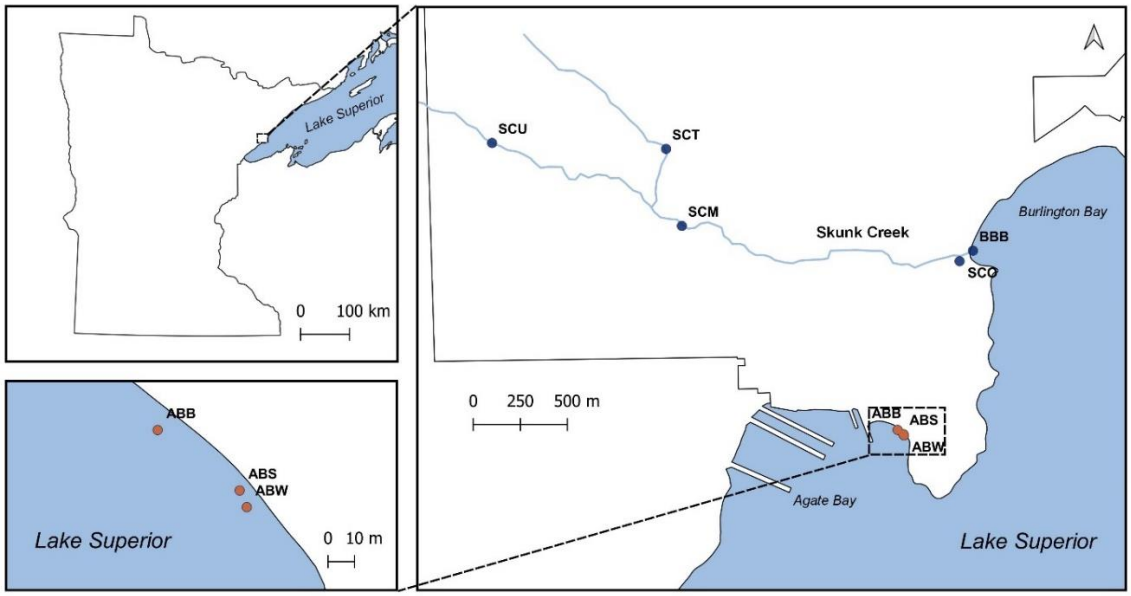


Figure 7. Map of the sampling locations in Two Harbors, MN. The upper left panel shows the geographic context of the study site along the North Shore of Lake Superior, the right panel shows all eight collection sites, and the lower left panel shows a close-up of the Agate Bay sites. Blue dots represent points in the Skunk Creek watershed while orange dots represent those in Agate Bay. Sites were sampled five times during summer 2019; three times during baseline conditions and twice within 24-hrs of a >1-inch rain event.

Laboratory Processing and Analysis

Water samples were immediately transported to the Natural Resources Research Institute (NRRI) where they were processed and prepared for molecular and water chemistry analyses. Fifty ml aliquots of raw sample were used for turbidity analysis performed with a Hach 2100 Series Laboratory Turbidimeter. Two-liter water samples were pre-filtered through a 70-micron mesh to remove large particles and debris. A 15 mL subsample of the mesh filtrate from each water sample was preserved in 0.5% HNO_3 for cation analysis (Ca, Mg, Fe, and K) using atomic adsorption spectrometry (AAS) (Varian AA240FS). Other subsamples were taken for culture-based *E. coli* analysis using the IDEXX

Colilert/QuantiTray system at 1x and 10x dilution using sterile phosphate-buffered saline (PBS) for a higher detection limit. The remaining mesh filtrate was then filtered through 5 um and 0.22 um nitrocellulose membranes (47 mm diameter) using vacuum filtration to collect microbial cells for molecular analysis. Filtered sample volumes were recorded and filters were placed in sterile Whirlpak bags, and frozen at -20°C until genomic materials were extracted. The filter heads and funnels were cleaned with a 10% bleach bath between samples. DNA extraction was performed using the MO BIO Laboratories PowerSoil DNA isolation kit using the instructions provided. Two milliliters of the filtrate were stored at 4°C for anion analysis (Cl^-, SO_4^-, NO_3^-) via Ion Chromatography (Dionex ICS 2000 with AS40 autosampler) .

Biomarker Analysis for Fecal Sources using qPCR

Quantitative polymerase chain reaction (qPCR) analysis determined the contribution of human and waterfowl fecal sources of *E. coli* to Skunk Creek and Lake Superior by using host-specific biomarkers (primers and probes) of human and avian sources for different qPCR assays (Table 5). For our human biomarkers, an established and sensitive human Bacteroides assay (HB) was used in conjunction with a new *Lachnospiraceae* genetic marker (Lachno3) shown to be highly specific for human fecal pollution (Feng et al. 2018). The avian biomarker GFD, which has been shown to be 100% avian-specific, occurs in geese, gulls, ducks and chickens (Green et al 2012). Standards for qPCR biomarkers were prepared using a wastewater effluent sample for the HB and Lachno3 biomarkers and Canada Geese (*Branta canadensis*) feces for the GFD biomarker. Canada

Geese droppings were collected from Harborview Park in Superior, WI and DNA was extracted with MO BIO Laboratories PowerFecal® DNA Isolation Kit. Conventional PCR and gel electrophoresis were performed to verify the size of target DNA segment for each marker. The amplified gene fragment was purified using an UltraClean 96 PCR Cleanup Kit and then used to prepare biomarker standards through serial dilution. Nine standards were used to construct qPCR standard curves for each assay and run alongside the samples (amplification efficiency 91.6-99.3%).

Table 5. Host-specific primers and probes for qPCR analyses of total bacteria (16S), and human (Lachno 3 and HB) and avian source (GFD) biomarkers of bacteria.

Marker	Forward Primer	Reverse Primer	Probe	Size	Annealing	Ref.
Lachno_3	5'-CAACGCGAAGAACCTTACCAAA-3'	5'-CCCAGAGTGCCCACCTTAAAT-3'	FAM-5'-CTCTGACCGGTCTTTAATCGGA-3'-MGB	186bps	64°C	Feng et al. 2018
HB	5'-ATCATGAGTTCACATGTCCG-3'	5'-CGTTACCCCGCCTACTATCTAATG-3'	6-FAM-5' TCCGGTAGACGATGGGGATGCGTT-TAMRA	86 bp	60°C	Feng et al. 2018
GFD	5'-TCGGCTGAGCACTCTAGGG-3'	5'-GCGTCTCTTTGTACATCCCA-3'	N/A	123 bp	56°C	Green et al. 2012
16S	5'-TCCTACGGGAGGCAGCAGT-3′	5' -GGA CTA CCA GGG TAT CTA ATC CTG TT- 3'	5'- CGT ATT ACC GCG GCT GCT GGC AC- 3'	466 bp	60°C	Nadkarni et al. 2020

The qPCR amplification program for all biomarkers included 1 cycle at 50°C for 2 min, followed by 1 cycle at 95°C for 10 min and then 40 cycles of 95°C for 15 s,

followed by 1 min at the biomarker's respective annealing temperature (64°C for Lachno3, 60°C for HB, 56°C for GFD and 60°C for 16S).

Statistical Analysis

Statistical analyses were performed using R software (version 3.5). T-tests, ANOVAs, correlation analysis and principal component analysis (PCA) were conducted to determine significant differences and help visualize the results. The threshold of significance for all analyses was $\alpha=0.05$. "Zero" values were set as one half of the detection limits for each instrument.

NGS and SourceTracker

DNA samples were sent to the University of Minnesota Genomics Center for partial sequencing of the V4 region of 16S rRNA gene using an Illumina MiSeq Version 3 Chemistry DNA sequencer. The returned sequence data was cleaned, aligned and clustered at 97% similarity to the SILVA rRNA database (v 1.32) using mothur MiSeq pipeline. These partial 16S rRNA gene sequence data were then imported into R and analyzed as described in Chapter 1. A permutational multivariate analysis of variance (PERMANOVA) was used to test the community differences between watersheds and site types (Anderson 2001). Redundancy analysis was performed at the order level of bacterial taxa using water quality data and centered log ratio (CLR) transformed 16S rRNA gene sequence data.

Library-dependent MST Using SourceTracker

Fecal taxonomic libraries were created using raw sequences of the V5-V6 region of the 16S rRNA gene from BioProject PRJNA377760 and BioProject PRJNA296920. The former sequences were collected from 209 fecal samples of 11 different animal species from northern Minnesota (Brown et al. 2017). The sequences were derived from a mixture of Illumina MiSeq and HiSeq sources and were assembled separately using Mothur (Schloss et al. 2009). Paired end sequences were trimmed to 150 bp and aligned with the SILVA ssRNA reference database (version 1.32, non-redundant). The latter sequences are three sources to model human fecal contamination were used. Human fecal samples were obtained from BioProject PRJNA296920, a study of the gut microbiomes of universal stool bank donors. Raw sequences of the V5-V6 region of the 16S rRNA gene were obtained from the Illumina Miseq platform (PE300). Twenty samples from distinct individuals were selected for use in this study. Paired end sequences were trimmed to 150 bp and aligned with the same SILVA ssRNA reference database.

In Brown et al.'s study, wastewater treatment plant influent was used as a standard for a human fecal signal. This work selected two datasets of wastewater influent to compare to the model based on human stool. Raw sequences of the V4-V5 region of the 16S rRNA gene were obtained from BioProject PRJNA597057, a five-year time series analysis of 73 wastewater treatment plants across the United States. Twenty samples of wastewater influent from treatment plants in Minnesota, Wisconsin, and Iowa were selected for this study and aligned using the same mothur protocol as other datasets. In addition, raw sequences of the V3-V4 region of 16S rRNA gene taken from municipal

wastewater samples in St. Louis County, MN were used as local representatives of wastewater influent.

Though the fecal libraries used in the SourceTracker analysis involved different hypervariable regions of the 16S rRNA gene, phylogenetic analyses of MiSeq V1–3, V3–4, V4, and V5-V6 datasets have shown that the taxonomic profiles of microbial communities have remarkably similar, and ultimately comparable, patterns at the phylum, family, and genus levels (Whon et al. 2018, Rajeev et al. 2020). In order to overcome the different sequence depths for the samples in this study with the fecal taxonomic library, read counts of OTUs in sample sets and the fecal libraries were agglomerated by genus taxa assignment before rarefication to a common depth. Sequence reads that could not be classified to the bacterial genus level were not included in this section of analysis.

16S rRNA gene sequence counts in the Agate Bay watershed were scaled to a minimum depth of 11,082 reads, covering 88% of unique bacterial genera in the sample set and 64% of unique genera in the fecal library. Sequence counts in the Skunk Creek watershed were scaled to a minimum depth of 12,353, covering 96% of unique bacterial genera in the sample set and 65% of unique genera in the fecal library. After unclassified genus level reads were dropped, the scaled read depths of Agate Bay and Skunk Creek sheds were reduced to 4,009 and 7,542 reads respectively.

SourceTracker was used to estimate proportions of fecal contamination in each watershed sample set (Knights et al. 2011). Initial model fit was improved by dropping training samples with incorrect majority classification, meaning samples that contained an unknown proportion larger than 25% and samples that did not contain a majority

classification. To assess the most appropriate source of human fecal signal for further use, the contribution of human fecal material to wastewater influent samples was modeled in a separate SourceTracker analysis. Sequences from the stool bank and animal fecal training sets were used as sources and both wastewater treatment plant datasets were used as sinks.

Results

Culture-based E. coli Abundance

Culturable *E. coli* abundance ranged from below the detection limit to 8.0×10^3 MPN/100 mL (Fig. 8a). *E. coli* levels in Skunk Creek's upstream site (SCU), had the largest average *E. coli* abundances with large variance (2682.1 MPN ± 3720.2 SD), followed by the Agate Bay Stormwater outfall (ABS) and Skunk Creek Outlet (SCO) sites. The Agate Bay Wastewater outfall (ABW) had the lowest average culturable *E. coli* counts (10.7 MPN ±11.1 SD) as the wastewater effluent was disinfected with chlorine during summer months (The State enforcement due to recreational water quality in Lake Superior). It is worth noting, that the *E. coli* counts for SCU only exceeded the state recreational standard (235 CFU/100 ml for an individual sample) during storm events, while *E. coli* abundance at the ABS and SCO sites exceeded this standard during both base flow and storm event samplings. The highest levels of culturable *E. coli* from the collection dates took place during the July storm event sampling. Considering all sites together, a paired, nested t-test demonstrated that *E. coli* abundance was significantly greater during storm event collection dates than for baseline collection dates (one-tailed p-value=0.028765, df=7, t=-2.27, Fig. 9). *E. coli* abundance was not significantly different between watersheds (t = -1.4047, df = 37, p-value = 0.1685), even though Skunk Creek had a slightly higher average.

Human and Avian Biomarkers

Human biomarkers (HB and Lachno3) were most prevalent at the Agate Bay sites (Fig. 8b, 8d). This trend was slightly more pronounced for the HB ($t = 3.9163$, $df = 30$, p-value = 0.0004803) than the Lachno3 biomarker ($t = 1.7845$, $df = 13.102$, p-value = 0.09751). A paired, nested t-test for all sampling locations showed average HB and Lachno3 copy numbers were similar between baseline and storm events (p=0.38 and 0.31 respectively). The avian biomarker (GFD) was most prominent at upstream Skunk Creek sites (SCT, SCM and SCU) and the Agate Bay stormwater outfall (ABS; Fig. 8c), although the avian biomarker relative abundance was similar in these two watersheds ($t = -1.1084$, $df = 30$, p-value = 0.2765). Storm events did not significantly impact abundance of the avian biomarker.

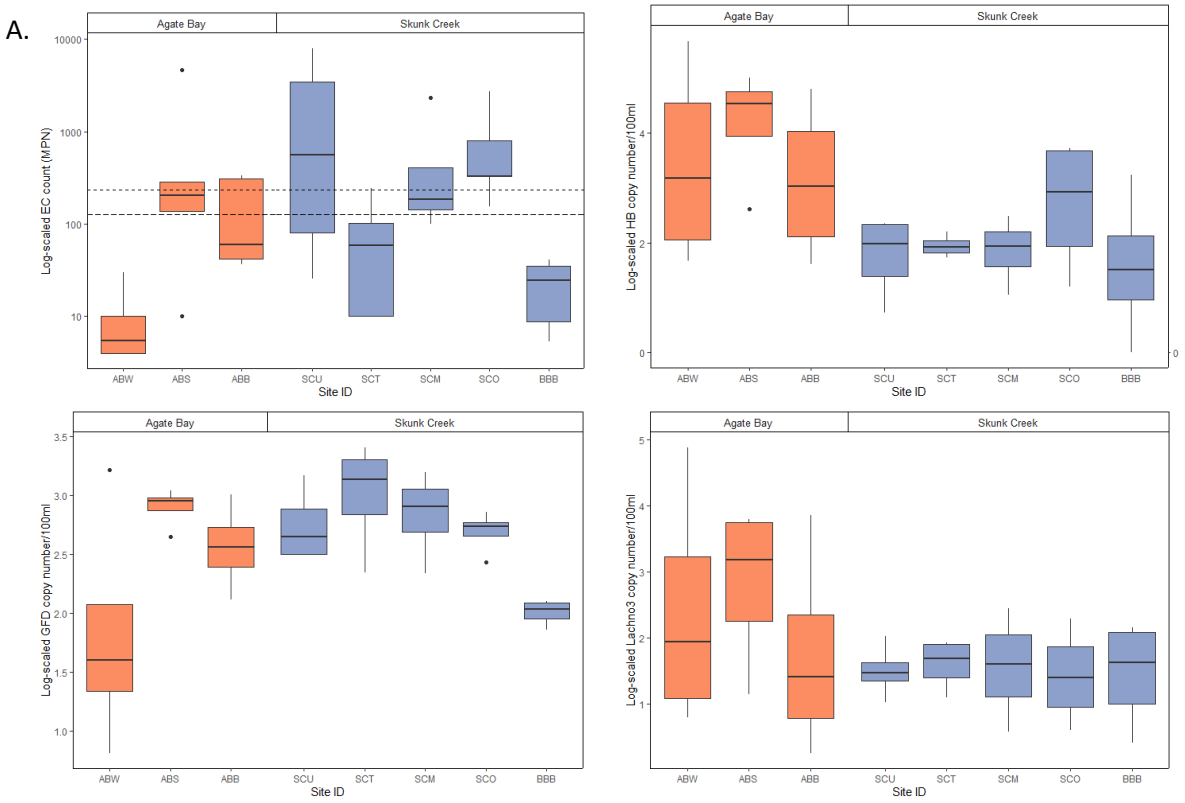


Figure 8. Log-scaled boxplots of (a) culturable *E. coli* (MPN), (b) human biomarker HB, (c) human biomarker Lachno3, and (d) avian biomarker GFD (copy numbers per 100 mL) in water from sites in Agate Bay and Skunk Creek during summer 2019. Horizontal lines in the first graph represent the Minnesota Lake Superior beach monitoring program sample standards (based on EPA guidelines). The top line is the single-sample threshold (235 MPN/100 ml) and below it is the geometric mean threshold (126 MPN/100 ml from at least five samples within a 30-day period).

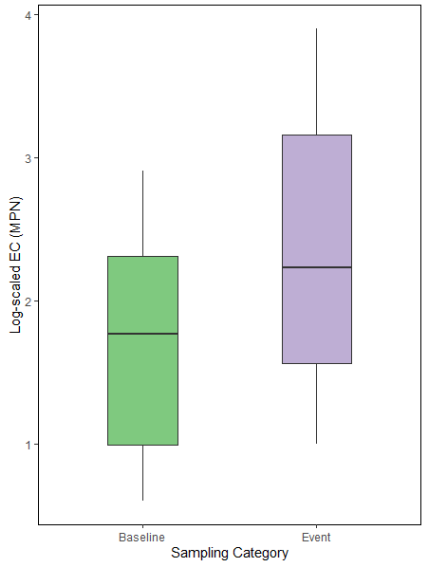


Figure 9. Log-scaled boxplots of culturable *E. coli* (MPN) by sampling category (baseline vs. storm-event) in water from Agate Bay and Skunk Creek sample sites in Two Harbors, MN.

Correlations Between Water Quality Parameters and Fecal Indicators

Cation analyses indicated that the Agate Bay Wastewater outfall (ABW) had higher magnesium concentrations than the other sampling locations ($p<0.05$; Fig. 10). Average calcium concentrations were greater at stream sites than bay sites and average iron concentrations were greatest at upstream sites (SCU, SCT and SCM). The Agate Bay Wastewater outfall had the highest concentrations of anions (Cl^-, SO_4^-, NO_3^-) distantly followed by the rest of the Agate Bay sites. Upstream sites had lower levels of dissolved oxygen, particularly the Skunk Creek Tributary. Principal component analysis revealed ABW to be the most distinct site based on measured water quality parameters (Fig. 10). A scatter plot matrix (SPLOM) with correlation coefficients and a Pearson's correlation test ($r = 0.53$) showed turbidity and *E. coli* counts were correlated ($t = 8.45$, $df = 37$, p

<0.001). There was a positive relationship between turbidity and *E. coli* counts for the Skunk Creek sites (y = 24.493x + 132.93, R^2 = 0.8286, p= 6.88043E-10).

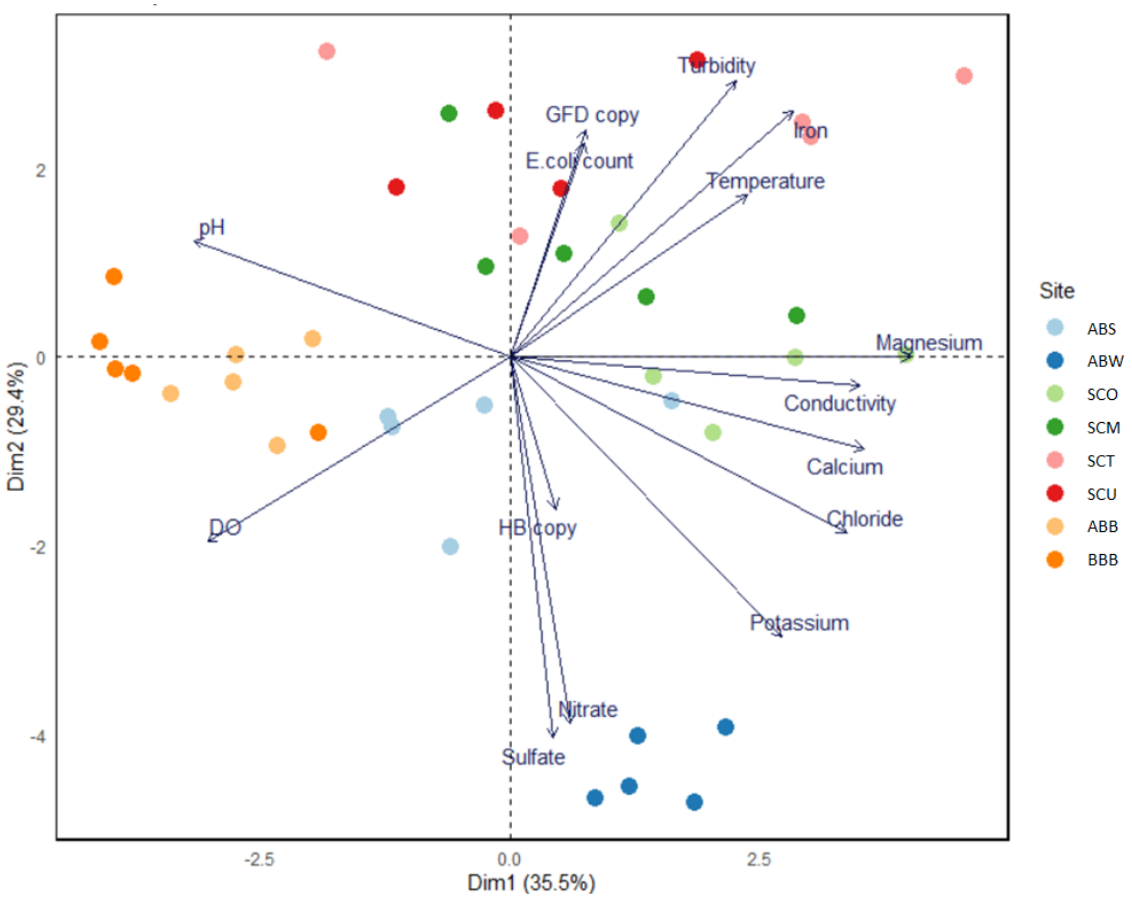


Figure 10. Principal component analysis of water quality data from Agate Bay, Burlington Bay, and Skunk Creek in Two Harbors, MN. Vectors represent water quality parameters and the colored shapes represent sampling sites (by site ID).

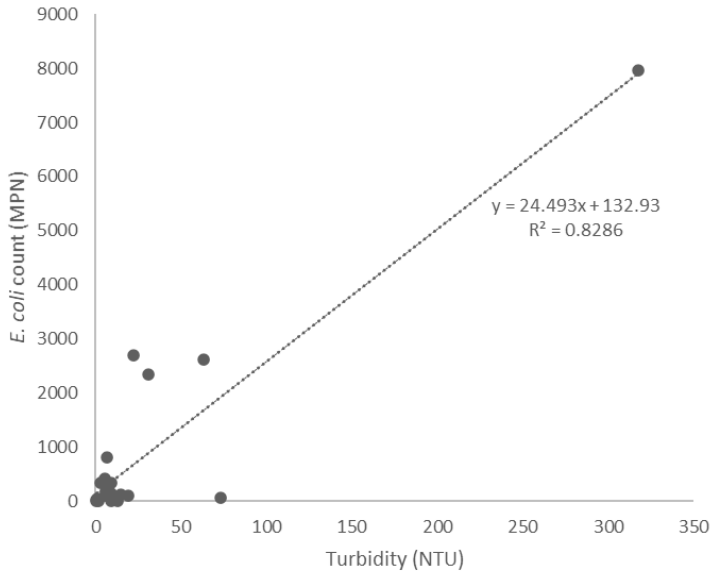


Figure 11. Linear regression of culturable *E. coli* (MPN) turbidity (NTU) for all Skunk Creek locations.

Bacterial Community Composition with NGS Sequencing

Between all environmental water samples, 69,425 OTUs were identified with the classification to 455 orders and 61 phyla. The bacterial communities of the samples were diversely comprised, but the most predominant order for all locations was Betaprotobacteriales (19.8±7.7%) from the phylum Proteobacteria. Among all samples, 6.1% of sequence reads could not be classified to an order.

Nearly all of the Lake Superior bay sites (ABB, ABS and BBB), had similar relative abundances of bacterial phyla (Fig. 12). Clustering of bacterial communities by site was confirmed by PCoA (Fig. 13) and PERMANOVA testing. Specifically, these analyses revealed significant clustering by watershed and site type ($p < 0.001$). These bay sites contained Chloroflexi and a higher proportion of Verromicrobia and Actinobacteria

compared to Skunk Creek stream sites. The Agate Bay Wastewater outfall had a greater proportion of Proteobacteria (the most of any of the project sites) and a notable presence of bacterial sequences from members of the Firmicutes and Espilonbacteria (unlike most of the other sites in Agate Bay). Like the Agate Bay Wastewater outfall site, the Agate Bay beach and stormwater outfall sites contained sequences from members of the Firmicutes bacterial phylum, while the Burlington Bay Beach site did not. Burlington Bay Beach had the largest relative abundance of sequences from the Cyanobacteria and Chloroflexi phyla of all the sampling sites. The Skunk Creek samples were distinguished by the presence of sequences from the Omnitrophicaeota phylum and a larger proportion of Patescibacteria phylum. The Skunk Creek sites also had the greatest amount of unclassified bacterial sequences.

The analysis of sequences from bacterial taxa to known harbor pathogens showed that water from the Agate Bay sites had a distinctive presence of sequences from members of the *Bacteriodes* genus, while the Skunk Creek Sites did not (Fig. 14). *Flavobacterium* was the pathogen-containing genus with the highest relative abundance of sequences for all sites except the wastewater outfall. The Agate Bay sites along with the Skunk Creek tributary had a notable presence of sequences from the *Pseudomonas* genus. Water from the Lake Superior sites all contained sequences from the *Mycobacterium* bacterial genus.

Redundancy analysis between the sequencing data and the water quality parameters showed that human biomarkers were associated with sequences from the Caulobacterales, Pirellulales, Rhodobacterales, Propionibacteriales, Leptospirales

bacterial orders (Fig. 15). Culturable *E. coli* and turbidity were correlated with bacterial 16S rRNA gene sequences from members of the Candidatus Falkowbacteria, Babeliales, and Alteromonadales orders.

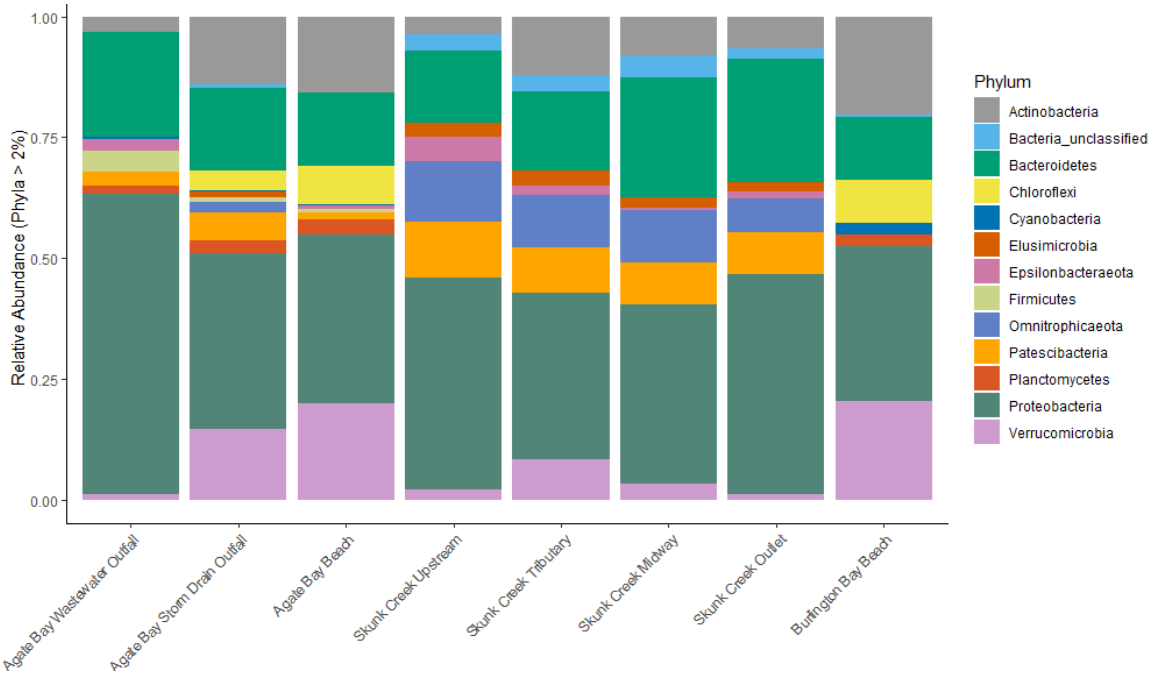


Figure 12. Relative abundance of phyla in bacterial communities (greater than 2%) in water from different sampling locations in Agate Bay, Burlington Bay, and Skunk Creek.

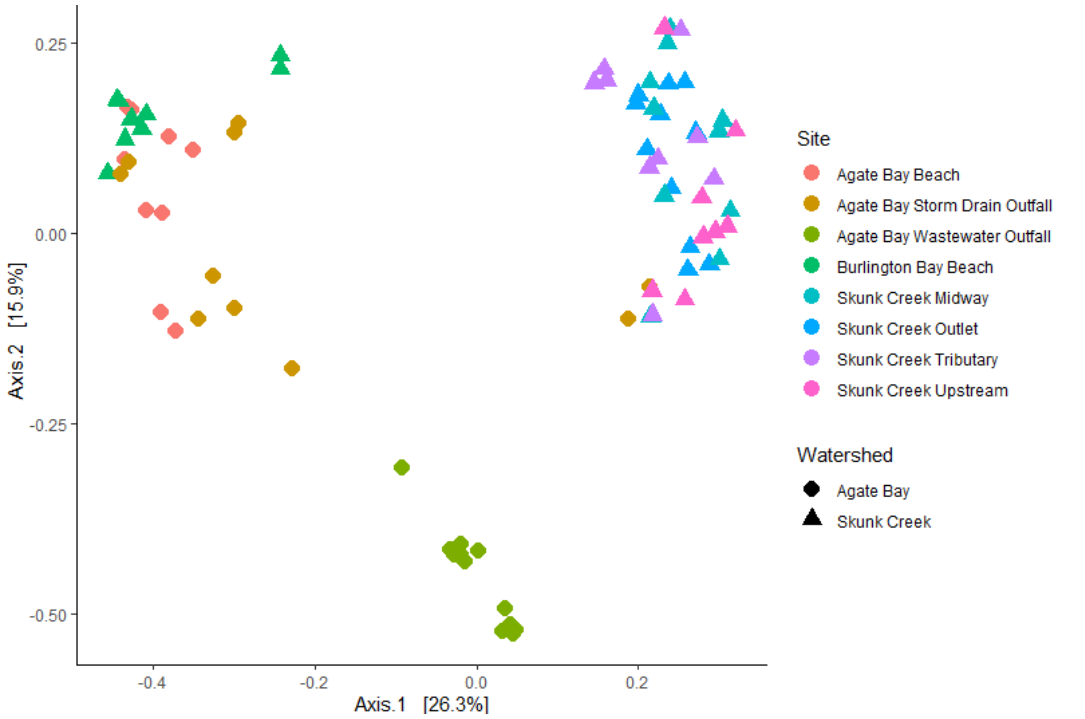


Figure 13. Principal coordinate analysis (PcoA) of partial 16S rDNA sequences from bacterial communities in Agate Bay, Burlington Bay, and Skunk Creek using the Bray-Curtis distance metric.

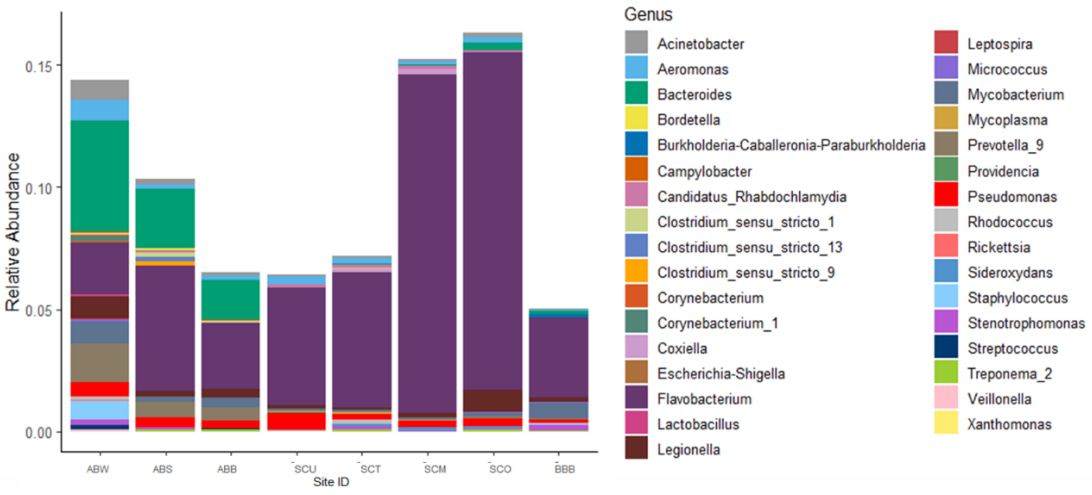


Figure 14. The relative abundance of pathogen-containing genera determined by partial 16S rDNA sequences (in read count) at different sample sites in Agate Bay (AB), Burlington Bay (BB), and Skunk Creek (SC). Skunk Creek sites are arranged upstream (left) to downstream (right).

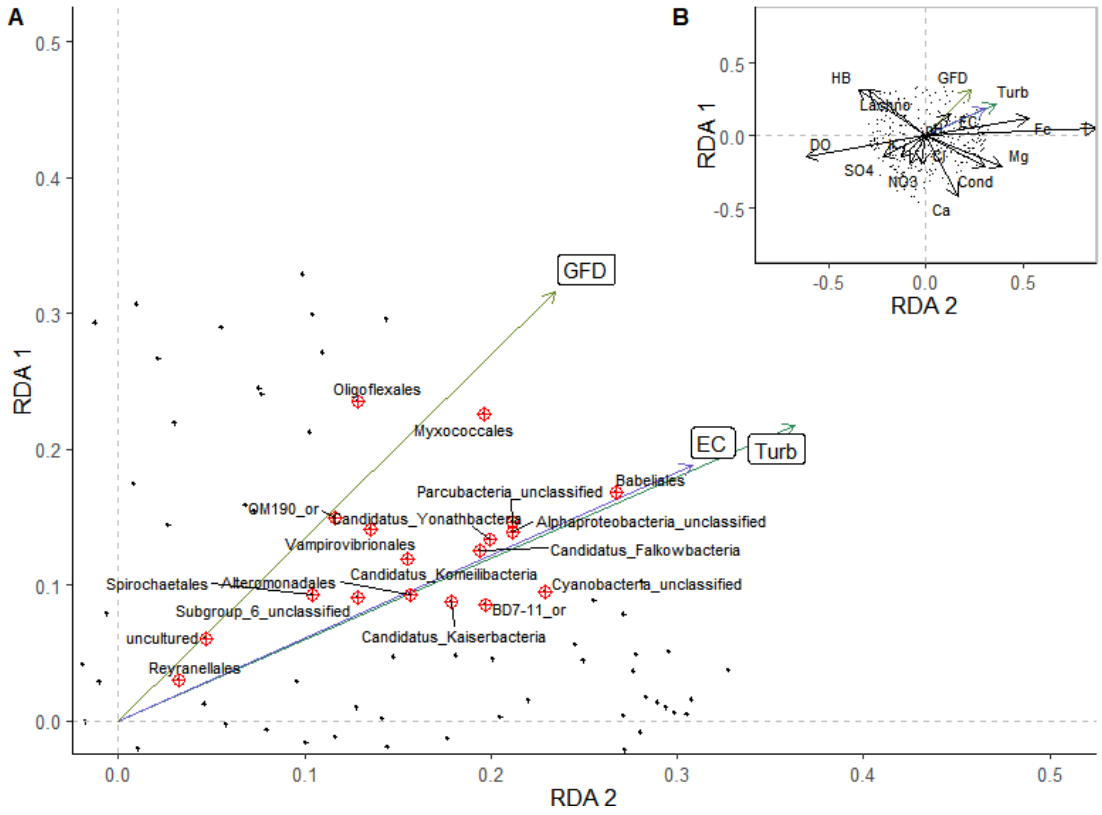


Figure 15. Redundancy (RDA) analysis of water quality parameters and partial 16S rDNA sequences from bacteria in water from Agate Bay, Burlington Bay, and Skunk Creek. The vectors represent water quality parameters and the data points represent Order-level bacterial taxa. For the water quality parameters, GFD represents the gene copy numbers for the avian biomarker, EC represents the level of culturable *E. coli* (MPN), and Turb represents turbidity (NTU). Panel A is a close-up of the upper-right quadrant of the entire RDA plot (panel B).

Library-dependent Fecal Signal in Watersheds

Fecal samples from 209 sources and 12 vertebrates were used to estimate the proportions of contamination from different sources in water samples from Agate Bay and Skunk Creek-Burlington Bay. This analysis indicated bacterial communities at sites in Agate Bay had a larger fecal bacteria contribution than Skunk Creek-Burlington Bay, which showed very little (almost no) fecal influence when compared to the fecal library (Fig. 16a and c). Of the fecal components in the Agate Bay samples, the largest identifiable

contribution was bacterial sequences from human bacterial sources, particularly during the June and early July collection dates (Fig. 15b). On those dates, a fecal signature from feline bacterial sources was the second most prevalent identifiable fecal bacterial source at the Agate Bay stormwater outfall site.

An internal comparison of the 16S rRNA gene sequences was performed using upstream sites as sources and downstream sites as sinks. This analysis showed that the Skunk Creek Upstream site had a greater influence on downstream sites during baseline events and the Skunk Creek Tributary had a larger influence during or after rain events (Fig. 17).

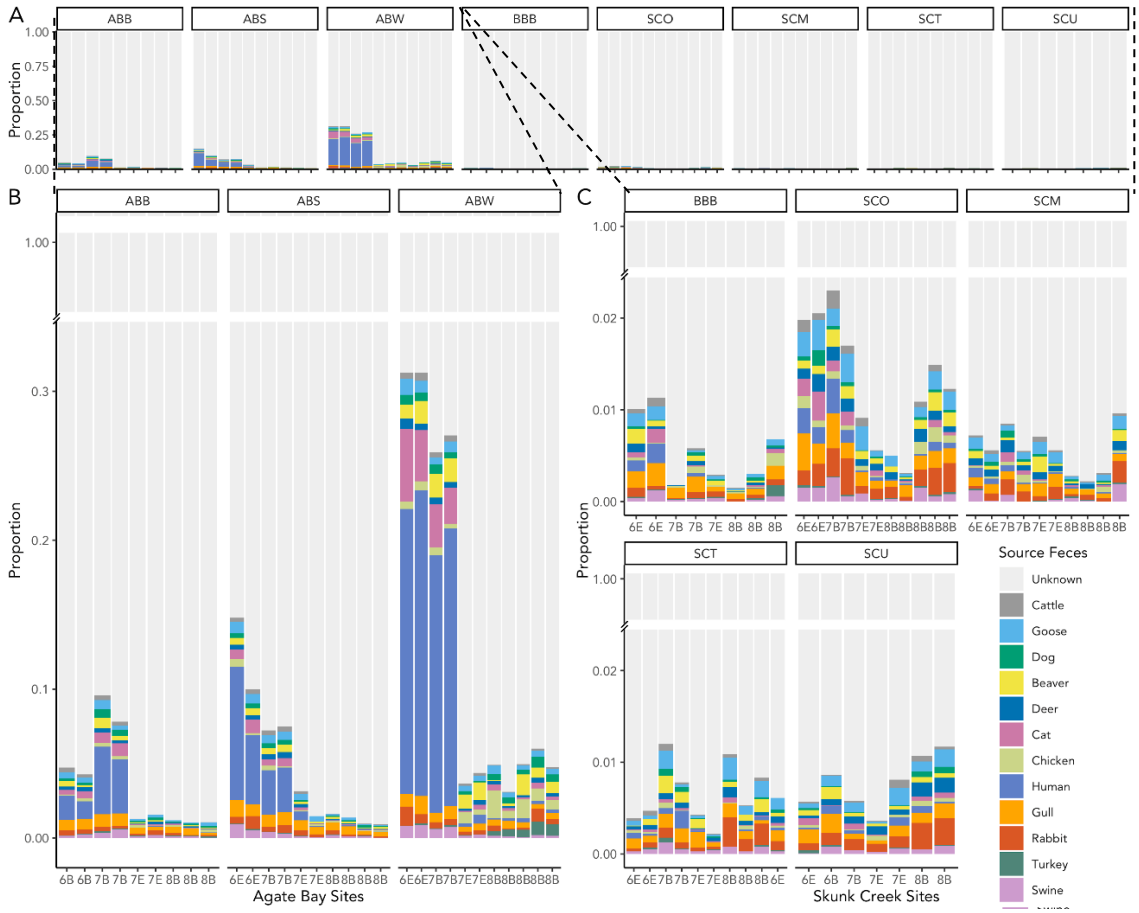


Figure 16. (a) Proportion of fecal contributions for all sampling locations in Two Harbors, MN (Agate Bay, Skunk Creek and Burlington Bay) generated through library-dependent SourceTracker analysis. (b) Close-up of fecal contributions in Agate Bay sampling locations with a y-axis range from 0.0-0.3 of the total proportion. (c) Close-up of fecal contributions in Skunk Creek sites with a y-axis range from 0.0-0.02 of the total proportion.

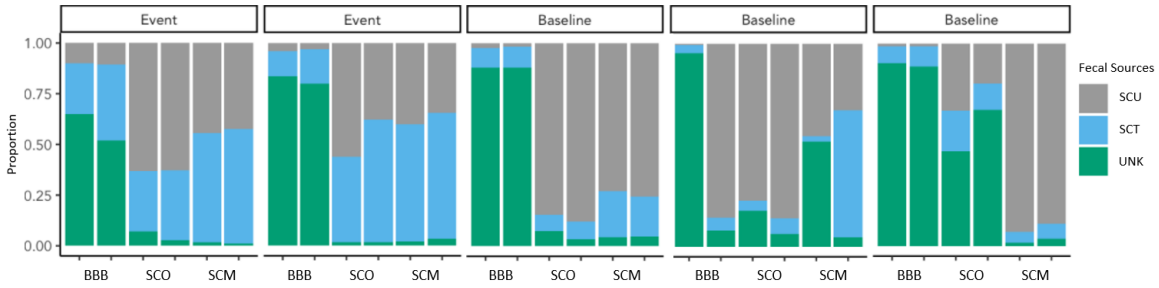


Figure 17. Proportions of fecal bacteria contributed from upstream sources from our internal SourceTracker analysis using the upstream Skunk Creek sites as sources and downstream locations as sinks. Water samples were collected during summer of 2019. Event sampling dates were 6/25 and 7/1 and baseline sampling dates were 7/1, 8/5, and 8/21. Upstream sites included Skunk Creek Upstream (SCU) and Skunk Creek Tributary (SCT) and downstream sites included Skunk Creek Midway (SCM), Skunk Creek Outlet (SCO) and Burlington Bay Beach (BBB). Green bars represented unknown sources (UNK), not linked to the source site inputs.

Discussion

The sources of fecal contamination were site-specific in this study. Both library-dependent and independent source tracking methods indicated a greater human influence at Agate Bay sites than Skunk Creek sites. In addition to the qPCR and SourceTracker results, the human influence on Agate Bay was further emphasized by sequencing data which showed *Bacteroides* and *Prevotella* (the two most prevalent Bacteroidetes genera in the human colon) as two pathogen-containing genera that distinguished Agate Bay from Skunk Creek sites (Fig. 13) (Ley 2016). The genetic information from treated wastewater (ABW) may account for a portion of the human genetic signature at the other Agate Bay sites. It does not, however, account for the exceedances in culturable *E. coli* or the days when the concentrations of human biomarkers in the Agate Bay stormwater outfall exceeded those from the wastewater outfall. While the specific inputs of the

human biomarkers are unknown, a sewer connection was detected and repaired upstream of the Agate Bay stormwater outfall soon after 2019 sampling ended. This connection may be partially responsible for the human inputs.

The higher contribution of avian sources at the Skunk Creek Tributary site compared to Agate Bay sites was likely influenced by the retention pond immediately upstream of the sample site. The retention pond appears to be hospitable to a variety of waterfowl, so an avian influence at this location is not surprising. The SourceTracker analysis, however, did not indicate a significant fecal influence on any of the Skunk Creek sites. This result supports the idea that avian fecal bacteria influence at the tributary site was a relatively small component of the largely non-fecal bacterial community.

If the spikes in culturable *E. coli* are not related to fecal inputs, one possible source could be naturalized *E. coli* populations within watersheds or streams (Ishii et al. 2006, Ksoll et al. 2007). Water turbidity at the Skunk Creek sites was positively correlated with culturable *E. coli* abundance. This relationship is not new. A study of roadside ditches (Falbo et al. 2013) showed that turbidity and concentrations/loadings of total suspended solids (TSS) were strong predictors of *E. coli* concentrations/loadings. Thus, the same mechanisms that erode and transport soils, sediment, or periphyton may also erode and transport *E. coli* communities living in or attached to these habitats.

There are a number of environmental factors that influence the occurrence and survival of *E. coli* in secondary, or extra-host, habitats (Petersen & Hubbart 2020). Soil

temperature is one factor, with *E. coli* growth increasing rapidly between 15° C and 37° C, suggesting that warmer months may see higher concentrations of *E. coli* in soils adjacent to streams (Ishii et al 2006). The greatest densities of soilborne *E. coli* in three Lake Superior watersheds were seen between June-October (Ishii et al. 2006), a timeframe that overlapped with the sampling period in this study. Dusek et al. (2018) also found increased prevalence of *E. coli* in forest habitats and areas in close proximity to forests. They suggested forests draining into streams could be contributing significant amounts of *E. coli* along with the suspended sediments in stormwater runoff. As the upstream portion of the Skunk Creek watershed is forested, this could be a contributor. Other fecal indicator bacteria, such as *Enterococcus* spp., have also been found to persistent in sediments and submerged aquatic vegetation (SAV) (Badgley et al. 2010). With this in mind, erosion and sediment control measures at key locations along the Skunk Creek could be beneficial to water quality in Burlington Bay.

The correlation between storm events and increased culturable *E. coli* indicated stormwater as a transport mechanism, particularly in the Skunk Creek watershed. Preliminary Two Harbors monitoring data from 2018 also showed a link between precipitation and *E. coli* concentrations because 12 of the 15 highest *E. coli* concentrations in the stream water occurred within 24 hours of a storm event (LSSWCD unpublished data).

Another key takeaway from this study is the importance of layering different analysis techniques to provide a fuller understanding of fecal sources. Based on the culturable *E. coli* indicator results, Skunk Creek's upstream site was the season's largest

source of fecal contamination. However, both the molecular biomarker and the SourceTracker approaches showed little genetic evidence of animal feces. The strong human influence at the Agate Bay stormwater outfall could have been dismissed as genetic material from the neighboring wastewater outfall, had there not been culturable *E. coli* data to distinguish the live sources unique to the stormwater outfall. Individually, each technique has strengths and limitations. Culture-based monitoring methods are user friendly and quantify live bacteria, but do not provide insight about the sources of contamination. Using qPCR with source-specific biomarkers provides information about potential sources, but can only evaluate one source at a time (though high-throughput, multiplexing technology like microfluidic qPCR could be a solution in that respect). Library-dependent methods like the SourceTracker program can analyze many sources simultaneously, but require proficiency in bioinformatics and programming languages. However, when all these techniques are used concurrently, they can complement each other to produce more complete and rigorous results.

Fecal contamination of recreational waters is a public health issue (as recreators risk exposure to pathogens), an economic issue (impacting tourism, real estate and healthcare) and an ecological issue (as nutrient-rich fecal inputs can lead to eutrophication and other downstream environmental issues). In addition to the disproportionately large contribution of nutrients from small Great Lakes tributaries, which tend to retain their inputs along coastlines, this emphasizes the importance of monitoring small streams like Skunk Creek.

In conclusion, pairing culturable *E. coli* measurements with library-dependent and independent source tracking methods provided a more complete picture of the potential sources of fecal contamination in Agate Bay and Skunk Creek. Levels of *E. coli* were correlated with turbidity and stormwater events, while human fecal sources appear to be site-specific and independent of storm events. Despite high abundances of culturable *E. coli*, library-dependent source tracking methods showed that the Skunk Creek watershed did not have a signature of fecal bacteria. This finding indicates that resuspended or eroded naturalized *E. coli* populations rather than *E. coli* cells of fecal origin may be responsible for the high *E. coli* abundances observed after storm events in Skunk Creek. Together, these results should be helpful for developing source-specific mitigation and management strategies for *E. coli* impairments in these two Lake Superior watersheds. The techniques and outcomes of this study may also be applicable to other small Great Lake tributary streams and recreational beaches experiencing similar stressors.

Research Implications

These two research applications of indicator bacteria highlight the benefits and limitations of their use in water quality monitoring. The benefit of simplifying pathogen monitoring to a few select bacteria with simple and well-established monitoring techniques is obvious. However, if the indicator bacteria are not effectively indicating the level of pathogens present, they are not serving their purpose. The ballast experiment showed us that the abundance of *E. coli* and *Enterococcus* sp. does not necessarily reflect the abundance of other pathogen containing genera. If we were to have relied solely on our indicator bacteria, we would have assumed that post-treatment regrowth of potential pathogens did not occur, which was not the case. In the microbial source tracking project, using culturable *E. coli* helped to distinguish living bacteria from killed bacteria in Agate Bay, but was not indicative of fecal contamination in Skunk Creek. These projects demonstrate that we can use indicator bacteria as a starting point, but we should look further into bacterial communities before making important decisions.

Both projects also reinforce the importance of layering multiple analysis techniques when using indicator bacteria. Culture-based methods, particularly the EPA approved IDEXX QuantiTray methodology, are a straightforward, user-friendly way to quantify live, culturable bacteria. However, they fail to account for viable but nonculturable bacteria which have the potential for reactivation and recovery. This can result in inaccurate quantification of bacterial regrowth and an overall underestimation of bacteria in a given sample. Molecular tools enable us to test for a wider range of organisms, but cannot necessarily distinguish between genetic material from living vs.

nonliving organisms. Molecular tools are also quickly evolving. Conventional qPCR, like the qPCR performed in these studies, can only quantify one target gene at a time. However, newer multiplexing techniques, like microfluidic qPCR are able to simultaneously quantify multiple pathogens alongside indicator bacteria. Overall, indicator bacteria are an imperfect monitoring tool, but their use can be strengthened by pairing traditional culture-based methods with complementary molecular techniques to provide a more comprehensive understanding of microbial communities in ballast treatment systems and recreational waterways.

References

Ahmed, W., Gyawali, P., & Toze, S. (2015). Quantitative PCR measurements of Escherichia coli including shiga toxin-producing *E. coli* (STEC) in animal feces and environmental waters. *Environmental Science and Technology*, *49*(5), 3084–3090. https://doi.org/10.1021/es505477n

Anderson, M.J. (2001). A new method for non-parametric multivariate analysis of variance. *Australian Ecology*, 26: 32-46.

Aw, T. G., Sivaganesan, M., Briggs, S., Dreelin, E., Aslan, A., Dorevitch, S., Haugland, R. (2019). Evaluation of multiple laboratory performance and variability in analysis of recreational freshwaters by a rapid Escherichia coli qPCR method (Draft Method C). *Water Research.* https://doi.org/10.1016/j.watres.2019.03.014

Badgley, B. D., Nayak, B.S., Harwood, V.J. (2010). The importance of sediment and submerged aquatic vegetation as potential habitats for persistent strains of enterococci in a subtropical watershed. *Water Research*, Volume 44, Issue 20, Pages 5857-5866,ISSN 0043-1354, https://doi.org/10.1016/j.watres.2010.07.005.

Badgley, B. D., Steele, M. K., Cappellin, C., Burger, J., Jian, J., Neher, T. P., Wagner, R. (2019). Fecal indicator dynamics at the watershed scale: Variable relationships with land use, season, and water chemistry. *Science of The Total Environment,* 697, 134113. https://doi.org/10.1016/j.scitotenv.2019.134113

Brown, C. M., Staley, C., Wang, P., Dalzell, B., Chun, C. L., & Sadowsky, M. J. (2017). A High-Throughput DNA-Sequencing Approach for Determining Sources of Fecal Bacteria in a Lake Superior Estuary. Retrieved from: https://doi.org/10.1021/acs.est.7b01353

Chung, W.S.F., Walker, A.W., Bosscher, D. (2020). Relative abundance of the *Prevotella* genus within the human gut microbiota of elderly volunteers determines the inter-individual responses to dietary supplementation with wheat bran arabinoxylan-oligosaccharides. *BMC Microbiol.* **20,** 283 https://doi.org/10.1186/s12866-020-01968-4

Cole, J. R., Wang, Q., Fish, J. A., Chai, B., McGarrell, D. M., Sun, Y., Tiedje, J. M. (2014). Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucleic Acids Research*. 42(D1), D633–D642. https://doi.org/10.1093/nar/gkt1244

de Bentzmann S, Plésiat P. (2011). The Pseudomonas aeruginosa opportunistic pathogen and human infections. *Environmental Microbiology*, Jul;13(7):1655-65. DOI: 10.1111/j.1462-2920.2011.02469.x

DeFlorio-Barker, S., Wing, C., Jones, R. M., & Dorevitch, S. (2018). Estimate of incidence and cost of recreational waterborne illness on United States surface waters. *Environmental Health: A Global Access Science Source*, *17*(1), 1–10. https://doi.org/10.1186/s12940-017-0347-9

Duluth Seaway Port Authority. (n.d.) Port Statistics. Retrieved from: http://www.duluthport.com/port-stats.php

Dusek, N., Hewitt, A. J., Schmidt, K. N., & Bergholz, P. W. (2018). Landscape-Scale Factors Affecting the Prevalence of Escherichia coli in Surface Soil Include Land Cover Type, Edge Interactions, and Soil pH. Retrieved from https://nhd.usgs.gov/

Falbo K, Schneider RL, Buckley DH, Walter MT, Bergholz PW, Buchanan BP. (2013). Roadside ditches as conduits of fecal indicator organisms and sediment: implications for water quality management. *The Journal of Environmental Management.* 128:1050–1059. https://doi.org/10.1016/j.jenvman.2013.05.021.

Falkinham, J. O. (2003). Factors influencing the chlorine susceptibility of Mycobacterium avium, Mycobacterium intracellulare, and Mycobacterium scrofulaceum. *Applied and Environmental Microbiology*, *69*(9), 5685–5689. https://doi.org/10.1128/AEM.69.9.5685-5689.2003

Field KG & Samadpour M (2007) Fecal source tracking, the indicator paradigm, and managing water quality. *Water Research.* 41: 3517–3538.

Feng, S., Bootsma, M., & Mclellan, S. L. (2018). Human-Associated Lachnospiraceae Genetic Markers Improve Detection of Fecal Pollution Sources in Urban Waters. Retrieved from https://doi.org/10

Fykse, E. M., Nilsen, T., Nielsen, A. D., Tryland, I., Delacroix, S., & Blatny, J. M. (2012). Real-time PCR and NASBA for rapid and sensitive detection of Vibrio cholerae in ballast water. *Marine Pollution Bulletin*, 64 (2), 200e206.

Gehring, T. M., & Santo, J. (2017). Potential for gulls to transport bacteria from human waste sites to beaches Science of the Total Environment Potential for gulls to transport bacteria from human waste sites to beaches. Science of the Total Environment, 615(October), 123–130. https://doi.org/10.1016/j.scitotenv.2017.09.232

Grabow, W.O.K. (1986) Indicator systems for assessment of the virological safety of treated drinking water. Water Science and Technology 18: 159-165.

Great Lakes Maritime Research Institute (GLMRI). (2012). Developing a Risk Assessment Tool to Predict the Risk of Accelerated Corrosion to Port Infrastructure-Final Report. Retrieved from: http://www.glmri.org/downloads/2011Reports/HicksandOster.pdf

Gregg, M., Rigby, G., & Hallegraeff, G. M. (2009). Review of two decades of progress in the development of management options for reducing or eradicating phytoplankton, zooplankton and bacteria in ship's ballast water. Aquatic Invasions, 4, 521–565. https://doi.org/10.3391/ai.2009.4.3.14

Grob, C., & Pollet, B.G., (2016). Regrowth in ship's ballast water tanks: think again! Marine Pollution Bulletin 109, 46–48. https://doi.org/10.1016/j.marpolbul.2016.04.061.

Hendrika J. De Lange, Donald P. Morris, Craig E. Williamson. (2003). Solar ultraviolet photodegradation of DOC may stimulate freshwater food webs. *Journal of Plankton Research*, Volume 25, Issue 1, Pages 111–117, https://doi.org/10.1093/plankt/25.1.111

Hess-Erga, O.-K., Blomvågnes-Bakke, B., & Vadstein, O. (2010). Recolonization by heterotrophic bacteria after UV irradiation or ozonation of seawater; a simulation of ballast water treatment. Water Research, 44(18), 5439–5449. https://doi.org/10.1016/J.WATRES.2010.06.059

Hess-Erga, O.-K., Moreno-Andrés, J., Enger, Ø., & Vadstein, O. (2019). Microorganisms in ballast water: Disinfection, community dynamics, and implications for management. *Science of The Total Environment*, 657, 704–716. https://doi.org/10.1016/j.scitotenv.2018.12.004

IMO. (n.d.). Ballast Water Management. Retrieved from: http://www.imo.org/en/OurWork/Environment/BallastWaterManagement/Pages/Default

IMO. (2019) Status of Treaties. Retrieved from: http://www.imo.org/en/About/Conventions/StatusOfConventions/Documents/StatusOfTreaties.pdf

Ishii, S., Ksoll, W. B., Hicks, R. E., & Sadowsky, M. J. (2006). Presence and growth of naturalized Escherichia coli in temperate soils from Lake Superior watersheds. *Applied and Environmental Microbiology*, 72(1), 612–621. https://doi.org/10.1128/AEM.72.1.612-621.2006

Johnson, L. K., M. B. Brown, E. A. Carruthers, J. A. Ferguson, P. E. Dombek, and M. J. Sadowsky. (2004). Sample size, library composition, and genotypic diversity among natural populations of escherichia coli from different animals influence accuracy of determining sources of fecal pollution. 2004. *Applied and Environmental Microbiololgy* **70**:4478-4485.

Juteau, P., Larocque, R., Rho, D., & Leduy, A. (1999). *Analysis of the relative abundance of different types of bacteria capable of toluene degradation in a compost biofilter*. Applied Microbiol Biotechnology. 52: 863-868

Kapoor, V., Pitkänen, T., Ryu, H., Elk, M., Wendell, D., & Domingo, J. W. S. (2015). Distribution of Human-Specific Bacteroidales and Fecal Indicator Bacteria in an Urban Watershed Impacted by Sewage Pollution, Determined Using RNA- and DNA-Based Quantitative PCR Assays. Retrieved from: https://doi.org/10.1128/AEM.02446-14

Kildare, B. J., Leutenegger, C. M., McSwain, B. S., Bambic, D. G., & Rajal, V. B. (2007). 16S rRNA-based assays for quantitative detection of universal, human-, cow-, and dog-specific fecal Bacteroidales: A Bayesian approach. *Water Research*, *41*(16), 3701–3715. https://doi.org/10.1016/J.WATRES.2007.06.037

Ksoll, W. B., Ishii, S., Sadowsky, M. J., & Hicks, R. E. (2007). Presence and sources of fecal coliform bacteria in epilithic periphyton communities of Lake Superior. Applied and Environmental Microbiology, 73(12), 3771–3778. https://doi.org/10.1128/AEM.02654-06

Knights, Dan, Justin Kuczynski, Emily S Charlson, Jesse Zaneveld, Michael C Mozer, Ronald G Collman, Frederic D Bushman, Rob Knight, and Scott T Kelley. (2011). "Bayesian Community-Wide Culture-Independent Microbial Source Tracking." *Nature Methods* 8 (9): 761–63.

Litchman, E. (2010). Invisible invaders: non-pathogenic invasive microbes in aquatic and terrestrial ecosystems. Ecology Letters, 13(12), 1560–1572. https://doi.org/10.1111/j.1461-0248.2010.01544.x

Ley, R. (2016). Prevotella in the gut: choose carefully. *Nat Rev Gastroenterol Hepatol* 13, 69–70 https://doi.org/10.1038/nrgastro.2016.4

LSRI – Lake Superior Research Institute. (2018). Duluth-Superior Harbor Water Experiment Summary – DRAFT.

Luo, L.-W., Wu, Y.-H., Yu, T., Wang, Y.-H., Chen, G.-Q., Tong, X., Hu, H.-Y. (2021). Evaluating method and potential risks of chlorine-resistant bacteria (CRB): A review. *Water Research*, *188*, 116474. https://doi.org/10.1016/j.watres.2020.116474

Marcarelli, A. M., Coble, A. A., Meingast, K. M., Kane, E. S., Brooks, C. N., Buffam, I., Stottlemyer, R. (2018). Of Small Streams and Great Lakes: Integrating Tributaries to Understand the Ecology and Biogeochemistry of Lake Superior. https://doi.org/10.1111/1752-1688.12695

Marshall S, Heath S, Henríquez V, Orrego C. (1998). Minimally invasive detection of Piscirickettsia salmonis in cultivated salmonids via the PCR. *Applied and Environmental Microbiology.* 64 (8): 3066–9. PMC 106817. PMID 9687475.

Matthews, Brian W. 1975. "Comparison of the Predicted and Observed Secondary Structure of T4 Phage Lysozyme." *Biochimica et Biophysica Acta (BBA)-Protein Structure* 405 (2): 442–51.

McGhee, J. J., Rawson, N., Bailey, B. A., Fernandez-Guerra, A., Sisk-Hackworth, L., & Kelley, S. T. (2020). Meta-SourceTracker: application of Bayesian source tracking to shotgun metagenomics. *PeerJ*, 8, e8783. https://doi.org/10.7717/peerj.8783

Minnesota Sea Grant (2017) Lake Superior's Non-Native Species. Retrieved from: http://www.seagrant.umn.edu/ais/superior_nonnatives

MPCA – Minnesota Pollution Control Agency (2018). Lake Superior South Watershed Total Maximum Daily Load Report Protecting and restoring Minnesota's North Shore resources. Retrieved from: www.pca.state.mn.uswww.tetratech.com

McMurdie, P. J., and Holmes, S. (2013). Phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PloS One* 8:e61217. Doi: 10.1371/journal.pone.0061217

Mooney, R. J., Stanley, E. H., Rosenthal, W. C., Esselman, P. C., Kendall, A. D., & Mcintyre, P. B. (2020.). Outsized nutrient contributions from small tributaries to a Great Lake. https://doi.org/10.1073/pnas.2001376117/-/DCSupplemental

Nadkarni, M. A., Martin, F. E., Jacques, N. A., and Hunter, N. (2002). Determination of bacterial load by real-time PCR using a broad-range (universal) probe and primers set. *Microbiology*, 148(1), 257–266. https://doi.org/10.1099/00221287-148-1-257

Ng, C., Le, T.-H., Goh, S. G., Liang, L., Kim, Y., Rose, J. B., & Gin Yew-Hoong, K. (2015). A Comparison of Microbial Water Quality and Diversity for Ballast and Tropical Harbor Waters. https://doi.org/10.1371/journal.pone.0143123

Okabe S, Okayama N, Savichtcheva O, Ito T. (2007). Quantification of host-specific *Bacteroides–Prevotella* 16S rRNA genetic markers for assessment of fecal pollution in freshwater. *Appl Microbiol Biotechnol* **74,** 890–901

Oksanen, J., Blanchet, G., Kindt, R. (2011) vegan: Community Ecology Package. R Package Version 2.0-2. https://cran.r-project.org/web/ packages/vegan/index.html (accessed 21 January 2017)

Pang, Y. C., Xi, J. Y., Xu, Y., Huo, Z. Y., & Hu, H. Y. (2016). Shifts of live bacterial community in secondary effluent by chlorine disinfection revealed by Miseq high-throughput sequencing combined with propidium monoazide treatment. *Applied Microbiology and Biotechnology*, *100*(14), 6435–6446. https://doi.org/10.1007/s00253-016-7452-5

Petersen, F., & Hubbart, J. A. (2020). Physical Factors Impacting the Survival and Occurrence of Escherichia coli in Secondary Habitats, 1–15.

Petersen, N. B., Madsen, T., Glaring, M. A., Dobbs, F. C., & Jørgensen, N. O. G. (2019). Ballast water treatment and bacteria: Analysis of bacterial activity and diversity after treatment of simulated ballast water by electrochlorination and UV exposure. *Science of The Total Environment*, 648, 408–421. https://doi.org/10.1016/j.scitotenv.2018.08.080

Porter, K.G., and Y.S. Feig. (1980). The use of DAPI for identifying and counting aquatic microflora. *Limnol. Oceanogr.* 25:943-948.

Rajeev, M., Sushmitha, T. J., Toleti, S. R., & Pandian, S. K. (2020). Sediment-associated bacterial community and predictive functionalities are influenced by choice of 16S ribosomal RNA hypervariable region(s): An amplicon-based diversity study. *Genomics*, 112(6), 4968–4979. https://doi.org/10.1016/j.ygeno.2020.09.006

Rueda, F. J., Fleenor, W. E., de Vicente, I. (2007). Pathways of river nutrients towards the euphotic zone in a deep-reservoir of small size: Uncertainty analysis. *Ecol. Modell,* 202, 345–361).

Ryu, H., Henson, M., Elk, M., Toledo-Hernandez, C., Griffith, J., Blackwood, D., Domingo, J. W. S. (2013). Development of Quantitative PCR Assays Targeting the 16S rRNA Genes of *Enterococcus spp.* and Their Application to the Identification of Enterococcus Species in Environmental Samples. https://doi.org/10.1128/AEM.02802-12

Saby, S., Sibille, I., Mathieu, L., Paquin, J. L., & Block, J. C. (1997). *Influence of Water Chlorination on the Counting of Bacteria with DAPI (4,6-Diamidino-2-Phenylindole) Downloaded from*. *Applied and Environmental Microbiology*, Vol. 63. Retrieved from http://aem.asm.org/

Scott, T. M., Rose, J. B., Jenkins, T. M., Farrah, S. R., & Lukasik, J. (2002). Microbial Source Tracking: Current Methodology and Future Directions. *Applied and Environmental Microbiology*, 68(12), 5796–5803. https://doi.org/10.1128/AEM.68.12.5796-5803.2002

Schloss, P.D., and Handelsman, J. (2005). Introducing dotur, a computer program for defining operational taxonomic units and estimating species richness. *Appl Environ Microbiol* 71: 1501– 1506.

Schloss, Patrick D, Sarah L Westcott, Thomas Ryabin, Justine R Hall, Martin Hartmann, Emily B Hollister, Ryan A Lesniewski. (2009). Introducing Mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. *Applied and Environmental Microbiology* 75 (23): 7537–41.

Soller, J. A., Schoen, M. E., Varghese, A., Ichida, A. M., Boehm, A. B., Eftim, S., Ravenscroft, J. E. (2014). Human health risk implications of multiple sources of faecal indicator bacteria in a recreational waterbody. *Water Research*, *66*, 254–264. https://doi.org/10.1016/j.watres.2014.08.026

Taylor, R. H., J. O. Falkinham III, C. D. Norton, and M. W. LeChevallier. 2000. Chlorine, chloramine, chlorine dioxide, and ozone susceptibility of Mycobacterium avium. Appl. *Environ. Microbiol.* 66:1702–1705.

Templar, H. A., Dila, D. K., Bootsma, M. J., Corsi, S. R., & McLellan, S. L. (2016). Quantification of human-associated fecal indicators reveal sewage from urban watersheds as a source of pollution to Lake Michigan. *Water Research*, *100*, 556–567. https://doi.org/10.1016/j.watres.2016.05.056

Tsolaki, E., & Diamadopoulos, E. (2010). Technologies for ballast water treatment: A review. *Journal of Chemical Technology and Biotechnology*, 85(1), 19–32. https://doi.org/10.1002/jctb.2276

Verma, D. K., & Rathore, G. (2015). New host record of five Flavobacterium species associated with tropical fresh water farmed fishes from North India. *Brazilian journal of*

microbiology : [publication of the Brazilian Society for Microbiology], *46*(4), 969–976. https://doi.org/10.1590/S1517-838246420131081

Waite, T., Kazumi, J., Lane, P., Farmer, L., Smith, S., Smith, S., Capo, T. (2003). Removal of natural populations of marine plankton by a large-scale ballast water treatment system. *Marine Ecology Progress Series*, 258, 51–63. https://doi.org/10.3354/meps258051

Wang, M., Ateia, M., Awfa, D., & Yoshimura, C. (2021). Regrowth of bacteria after light-based disinfection d What we know and where we go from here. *Chemosphere*, *268*, 128850. https://doi.org/10.1016/j.chemosphere.2020.128850

Whon, T. W., Chung, W. H., Lim, M. Y., Song, E. J., Kim, P. S., Hyun, D. W., Nam, Y. Do. (2018). Data Descriptor: The effects of sequencing platforms on phylogenetic resolution in 16 S rRNA gene profiling of human feces. *Scientific Data*, 5, 1–15. https://doi.org/10.1038/sdata.2018.68

United States Census Bereau. (2018). American Community Survey Demographic and Housing Estimates https://data.census.gov/

USEPA (n.d.) Water Monitoring & Assessment - Fecal Bacteria. https://archive.epa.gov/water/archive/web/html/vms511.html

USEPA - Office of Wastewater Management. (2011). Ballast Water Self-Monitoring. Retrieved from https://www3.epa.gov/npdes/pubs/vgp_ballast_water.pdf

USEPA. (2014). EPA Draft - Method C: Escherichia coli in Water by TaqMan Quantitative Polymerase Chain Reaction (qPCR) Assay.

Zweifel UL, Hagstrom A. (1995). Total counts of marine bacteria include a large fraction of non-nucleoid-containing bacteria (ghosts). *Appl Environ Microbiol*, 61(6):2180-5. doi: 10.1128/aem.61.6.2180-2185.1995. PMID: 16535043; PMCID: PMC1388461.

Appendix – Supplementary Material

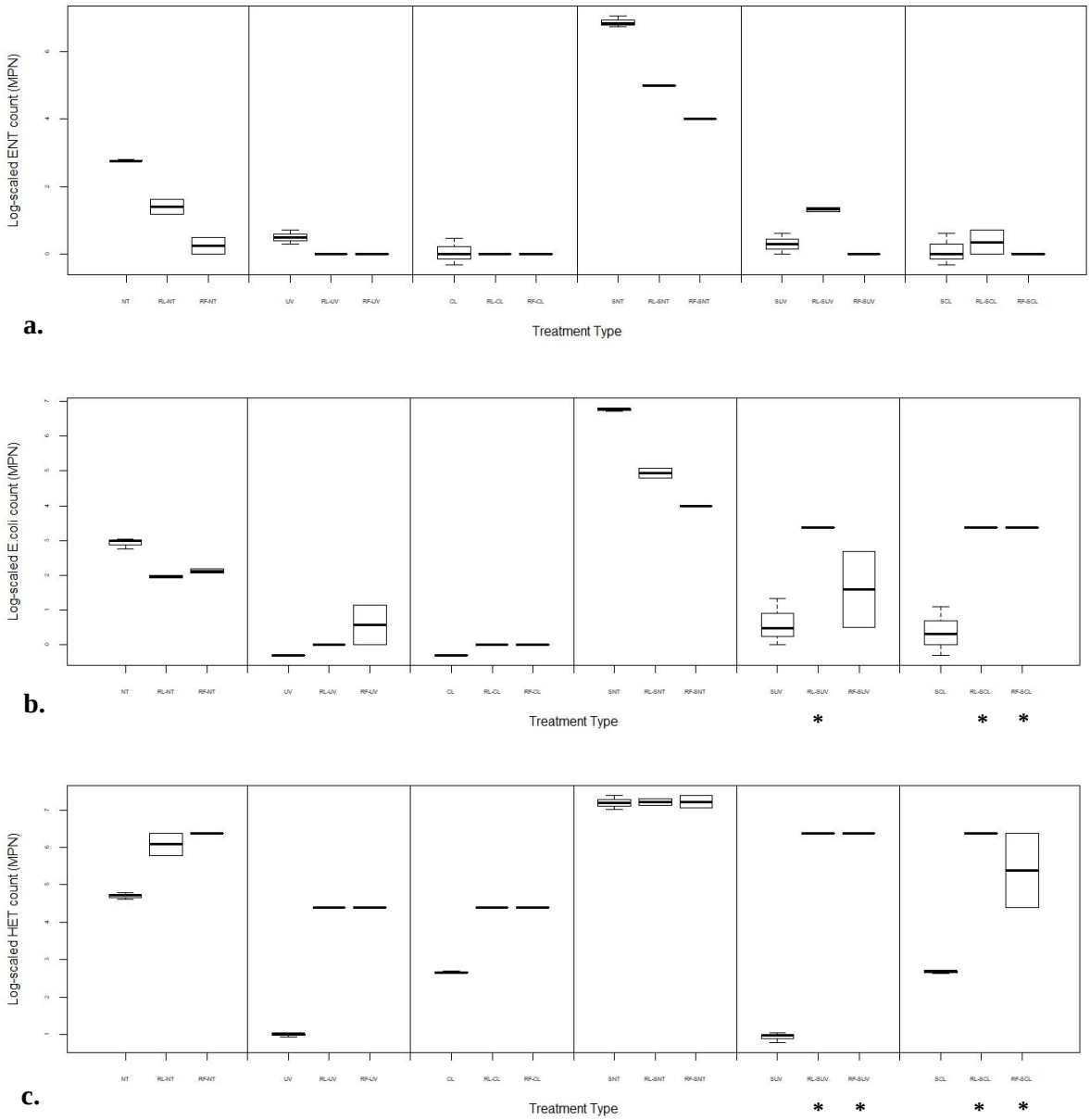
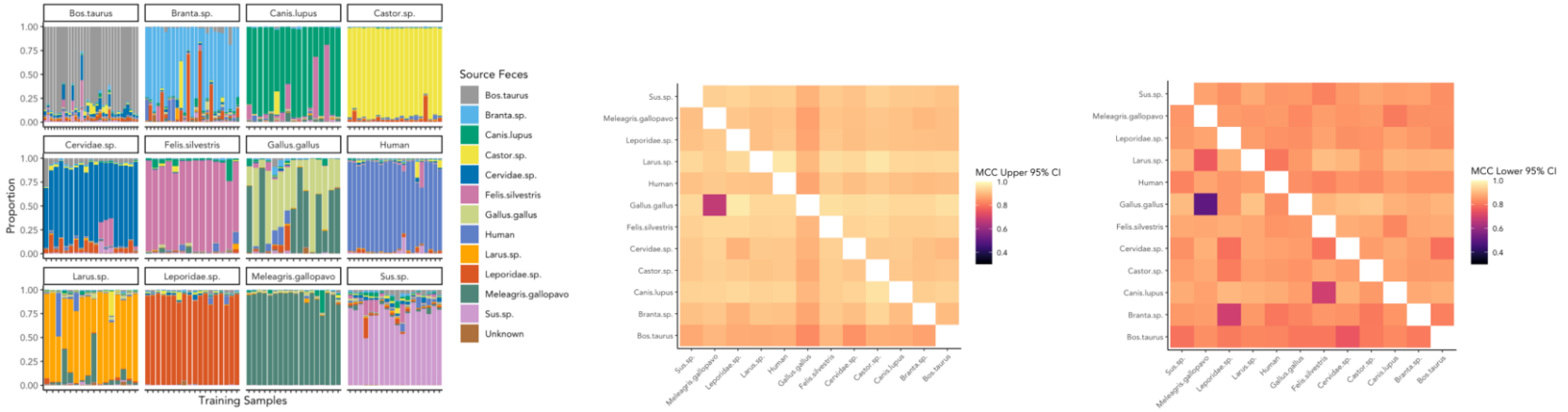


Figure S1. Abundance of target bacteria determined by culture-based analysis. (a) Log-scaled abundance of *Enterococcus* sp. (MPN), (b) log-scaled abundance of *E. coli* (MPN), and (c) log-scaled abundance of total heterotrophic bacteria (MPN). An asterisk below the treatment type indicates that the sample reached the maximum limit of detection for its dilution factor and is therefore an underestimation.

Table S1. Water quality measurements from sites in Agate Bay and Skunk Creek during summer 2019. Each value is the average of duplicate samples followed by the standard deviation (SD), with the exception of field measurements.

Site No.	Site Name	Sampling Date	Category	Temp (°C)	DO (mg/L)	pH	Conductivity (us)	Turbidity (NTU) ± SD	Fe^{2+} (ppm) ± SD	Mg^{2+} (ppm) ± SD	Ca^{2+} (ppm) SD	K^{+} (ppm) ± SD	Cl- (ppm) ± SD	SO4- (ppm) ± SD	NO3- (ppm) ± SD
2	Agate Bay Storm Drain Outfall	06/25/19	Event	7.4	10.93	7.88	321.9	3.60 ± 0.00	0.08 ± 0.02	0.66 ± 0.04	7.80 ± 0.24	0.38 ± 0.07	11.80 ± 0.43	10.20 ± 0.81	6.47 ± 0.15
3	Agate Bay Wastewater Outfall	06/25/19	Event	9.5	11.06	7.34	323.9	2.15 ± 0.00	0.05 ± 0.02	1.10 ± 0.08	10.77 ± 1.91	1.19 ± 0.17	47.65 ± 4.06	34.10 ± 4.49	25.51 ± 3.91
4	Skunk Creek Outlet	06/25/19	Event	12.3	9.36	7.85	427	5.75 ± 0.00	0.37 ± 0.05	1.54 ± 0.10	15.67 ± 1.23	0.34 ± 0.00	28.28 ± 3.65	4.31 ± 0.70	1.15 ± 0.04
7	Skunk Creek Midway	06/25/19	Event	13.2	9.56	8.05	263.1	6.00 ± 0.00	0.18 ± 0.05	1.05 ± 0.21	8.05 ± 2.40	0.19 ± 0.03	10.31 ± 0.37	2.47 ± 0.13	0.41 ± 0.02
8	Skunk Creek Tributary	06/25/19	Event	15.4	7.6	7.9	236.7	9.30 ± 0.00	0.57 ± 0.01	1.00 ± 0.09	5.87 ± 0.56	0.13 ± 0.00	15.42 ± 0.14	2.17 ± 0.03	0.38 ± 0.01
9	Skunk Creek Transition	06/25/19	Event	12.4	9.56	7.97	191.4	14.40 ± 0.00	0.43 ± 0.01	0.68 ± 0.03	3.70 ± 1.00	0.14 ± 0.01	3.26 ± 0.68	2.01 ± 0.07	0.42 ± 0.02
11	Agate Bay Beach	06/25/19	Event	6.2	12.2	7.95	124.4	2.46 ± 0.00	0.08 ± 0.02	0.32 ± 0.02	4.87 ± 0.97	0.12 ± 0.01	3.32 ± 1.30	4.49 ± 0.02	2.44 ± 0.01
12	Burlington Bay Beach	06/25/19	Event	6	12.25	8.08	144.6	1.15 ± 0.00	0.10 ± 0.03	0.81 ± 0.44	10.73 ± 6.98	0.12 ± 0.00	7.14 ± 0.68	3.65 ± 0.44	1.40 ± 0.28
2	Agate Bay Storm Drain Outfall	07/01/19	Baseline	7.5	10.83	8	187	3.98 ± 1.88	0.14 ± 0.08	0.59 ± 0.31	6.27 ± 5.15	0.19 ± 0.09	6.94 ± 0.36	8.30 ± 0.15	4.94 ± 0.09
3	Agate Bay Wastewater Outfall	07/01/19	Baseline	11.05	10.39	7.38	505	2.78 ± 0.04	0.10 ± 0.01	1.30 ± 0.03	13.86 ± 0.82	1.74 ± 0.02	52.82 ± 3.04	50.04 ± 1.50	36.66 ± 0.56
4	Skunk Creek Outlet	07/01/19	Baseline	13.9	6.68	7.54	529	6.40 ± 3.25	0.85 ± 0.04	1.75 ± 0.21	19.45 ± 1.55	0.39 ± 0.06	47.75 ± 4.24	5.59 ± 0.51	2.77 ± 0.38
7	Skunk Creek Midway	07/01/19	Baseline	15.7	8.73	7.88	361.1	5.10 ± 0.57	0.14 ± 0.03	1.16 ± 0.50	9.64 ± 5.02	0.19 ± 0.11	15.92 ± 1.17	3.15 ± 0.26	0.55 ± 0.02
8	Skunk Creek Tributary	07/01/19	Baseline	14.6	2.3	7.15	515	12.88 ± 6.68	1.19 ± 0.57	1.57 ± 0.50	13.21 ± 5.83	0.11 ± 0.06	20.16 ± 0.78	0.85 ± 0.00	0.47 ± 0.01
9	Skunk Creek Transition	07/01/19	Baseline	14.2	8.15	7.82	313.5	12.40 ± 1.27	0.43 ± 0.06	1.40 ± 0.25	12.39 ± 2.82	0.22 ± 0.05	3.38 ± 0.02	1.03 ± 0.03	0.50 ± 0.01
11	Agate Bay Beach	07/01/19	Baseline	6.2	13.04	8.04	109.1	3.42 ± 0.12	0.08 ± 0.03	0.42 ± 0.01	5.69 ± 0.17	0.18 ± 0.04	5.43 ± 0.10	6.94 ± 0.08	4.36 ± 0.08
12	Burlington Bay Beach	07/01/19	Baseline	7.2	13.07	8.28	103	1.36 ± 0.77	0.07 ± 0.01	0.24 ± 0.13	4.37 ± 1.07	0.06 ± 0.00	2.49 ± 0.36	3.16 ± 0.06	1.44 ± 0.01
2	Agate Bay Storm Drain Outfall	07/29/19	Event	13.3	10.32	8.11	532	6.65 ± 0.07	0.21 ± 0.01	1.47 ± 0.29	24.42 ± 6.27	0.41 ± 0.11	13.52 ± 0.34	10.04 ± 0.25	6.97 ± 0.07
3	Agate Bay Wastewater Outfall	07/29/19	Event	14.8	11.63	7.42	395	0.74 ± 0.00	0.11 ± 0.01	0.94 ± 0.17	12.39 ± 0.00	1.32 ± 0.02	45.44 ± 3.28	34.34 ± 0.74	52.49 ± 2.26
4	Skunk Creek Outlet	07/29/19	Event	18.1	9.99	8.13	291.4	22.25 ± 0.35	0.45 ± 0.05	1.09 ± 0.15	11.08 ± 1.38	0.27 ± 0.02	25.06 ± 7.78	3.15 ± 0.01	1.13 ± 0.15
7	Skunk Creek Midway	07/29/19	Event	18.7	9.91	8.2	211	30.75 ± 0.35	0.46 ± 0.05	0.32 ± 0.27	5.15 ± 1.83	0.13 ± 0.05	9.99 ± 0.11	2.19 ± 0.04	0.58 ± 0.03
8	Skunk Creek Tributary	07/29/19	Event	20.8	8.14	8.69	197.8	6.60 ± 0.14	0.30 ± 0.01	0.77 ± 0.06	0.71 ± 1.01	0.10 ± 0.01	11.28 ± 0.04	1.80 ± 0.02	0.05 ± 0.00
9	Skunk Creek Transition	07/29/19	Event	16.7	10.26	7.75	153.3	63.40 ± 1.56	0.85 ± 0.03	0.52 ± 0.33	5.21 ± 1.21	0.18 ± 0.01	4.45 ± 0.07	2.78 ± 0.17	0.67 ± 0.01
11	Agate Bay Beach	07/29/19	Event	12	13.38	8.2	113.3	1.48 ± 0.00	0.15 ± 0.02	0.31 ± 0.00	4.75 ± 0.09	0.10 ± 0.00	3.59 ± 0.89	4.94 ± 0.21	3.16 ± 0.15
12	Burlington Bay Beach	07/29/19	Event	12.3	12.69	8.38	101.6	1.19 ± 0.25	0.08 ± 0.00	0.25 ± 0.08	4.40 ± 0.70	0.04 ± 0.02	1.70 ± 0.00	3.71 ± 0.00	1.53 ± 0.00
2	Agate Bay Storm Drain Outfall	08/05/19	Baseline	11.9	8.52	8.23	576	1.53 ± 0.81	0.05 ± 0.02	0.99 ± 0.11	9.29 ± 0.22	0.22 ± 0.02	6.18 ± 1.88	6.61 ± 1.48	5.09 ± 1.41
3	Agate Bay Wastewater Outfall	08/05/19	Baseline	16.2	9.89	7.27	412.9	0.49 ± 0.01	0.06 ± 0.02	1.49 ± 0.03	13.63 ± 0.63	1.87 ± 0.04	54.89 ± 0.72	47.59 ± 1.23	87.58 ± 2.17
4	Skunk Creek Outlet	08/05/19	Baseline	17	5.67	7.5	678	9.00 ± 0.71	1.64 ± 0.07	2.66 ± 0.03	31.65 ± 0.11	0.57 ± 0.02	47.53 ± 2.47	10.65 ± 0.21	3.16 ± 0.05
7	Skunk Creek Midway	08/05/19	Baseline	18.5	7	7.84	377.2	5.80 ± 2.40	0.10 ± 0.00	1.77 ± 0.03	15.67 ± 0.37	0.30 ± 0.01	17.76 ± 0.03	3.81 ± 0.11	0.71 ± 0.03
8	Skunk Creek Tributary	08/05/19	Event	18	3.24	7.58	351.1	18.75 ± 6.72	1.80 ± 2.37	2.31 ± 0.08	19.94 ± 0.27	0.25 ± 0.01	10.86 ± 0.18	2.18 ± 0.03	0.70 ± 0.02
9	Skunk Creek Transition	08/05/19	Event	16.9	7.7	7.87	302.3	317.50 ± 0.71	3.38 ± 1.44	1.46 ± 0.41	11.28 ± 4.34	0.29 ± 0.14	3.52 ± 0.24	2.07 ± 0.03	1.36 ± 0.18
11	Agate Bay Beach	08/05/19	Baseline	16.3	10.82	8.52	1021	0.53 ± 0.21	0.05 ± 0.01	0.35 ± 0.01	5.28 ± 0.35	0.08 ± 0.00	1.94 ± 0.20	4.09 ± 0.25	1.95 ± 0.11
12	Burlington Bay Beach	08/05/19	Baseline	15.2	11.28	8.54	101.4	0.50 ± 0.11	0.03 ± 0.01	0.37 ± 0.02	3.92 ± 0.04	0.07 ± 0.00	1.98 ± 0.46	3.89 ± 0.36	1.67 ± 0.06
2	Agate Bay Storm Drain Outfall	08/21/19	Baseline	8.8	11.32	7.48	187	1.24 ± 0.13	0.10 ± 0.00	0.50 ± 0.00	5.99 ± 0.25	0.23 ± 0.01	6.08 ± 0.64	7.26 ± 1.22	6.77 ± 0.78
3	Agate Bay Wastewater Outfall	08/21/19	Baseline	10	11.49	7.31	241	0.84 ± 0.03	0.16 ± 0.05	1.10 ± 0.17	12.69 ± 0.48	1.86 ± 0.06	55.87 ± 0.95	52.56 ± 1.08	71.52 ± 2.79
4	Skunk Creek Outlet	08/21/19	Baseline	14.7	8.4	7.5	500	2.62 ± 0.03	0.53 ± 0.14	1.46 ± 0.12	15.12 ± 0.44	0.41 ± 0.10	47.64 ± 0.59	6.43 ± 0.21	2.50 ± 1.21
7	Skunk Creek Midway	08/21/19	Baseline	14.7	5.87	7.05	454.3	9.00 ± 4.24	0.31 ± 0.06	2.21 ± 1.17	22.42 ± 14.50	0.20 ± 0.01	44.21 ± 0.20	3.80 ± 0.15	0.80 ± 0.03
8	Skunk Creek Tributary	08/21/19	Baseline	15.3	1.48	6.88	636	73.20 ± 3.11	7.36 ± 0.07	2.10 ± 0.15	17.70 ± 0.68	0.16 ± 0.02	12.32 ± 0.06	0.81 ± 0.01	1.58 ± 0.21
11	Agate Bay Beach	08/21/19	Baseline	6.3	12.64	7.66	105.3	0.90 ± 0.03	0.10 ± 0.01	0.30 ± 0.08	3.73 ± 0.12	0.07 ± 0.03	2.24 ± 0.08	3.85 ± 0.01	2.39 ± 0.11
12	Burlington Bay Beach	08/21/19	Baseline	6	13.19	8.14	102.6	0.51 ± 0.08	0.11 ± 0.02	0.37 ± 0.03	2.05 ± 2.89	0.06 ± 0.00	1.74 ± 0.02	3.62 ± 0.30	1.65 ± 0.06

A.



B.

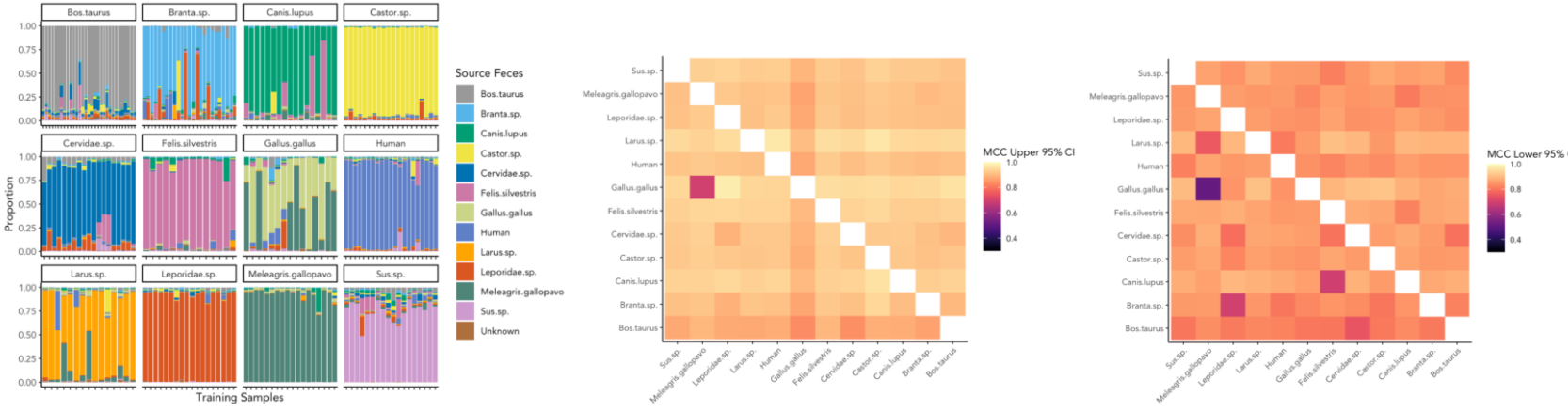


Figure S2. Model validation for the SourceTracker analysis. Training samples and Matthew's Correlation Coeficients (MCC) for a.) Agate Bay and b.) Skunk Creek samples.